

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/11, 15/10, C07K 14/47, C12P 21/00, C12Q 1/68, C07K 16/18, G06F 17/30, 17/50</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 99/53051</b> <b>(43) International Publication Date:</b> 21 October 1999 (21.10.99)
<b>(21) International Application Number:</b> PCT/IB99/00712 <b>(22) International Filing Date:</b> 9 April 1999 (09.04.99)  <b>(30) Priority Data:</b> 09/057,719 9 April 1998 (09.04.98) US 09/069,047 28 April 1998 (28.04.98) US  <b>(71) Applicant (for all designated States except US):</b> GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> DUMAS MILNE EDWARDS, Jean-Baptiste [FR/FR]; 8, rue Grégoire-de-Tours, F-75006 Paris (FR). DUCLERT, Aymeric [FR/FR]; 6 ter, rue Victorine, F-94100 Saint-Maur (FR). GIORDANO, Jean-Yves [FR/FR]; 12, rue Duhesme, F-75018 Paris (FR).  <b>(74) Agents:</b> MARTIN, Jean-Jacques et al.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).		<b>(81) Designated States:</b> AU, CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> 5' ESTS AND ENCODED HUMAN PROTEINS  <b>(57) Abstract</b>  The sequences of 5' ESTs derived from mRNAs encoding secreted proteins are disclosed. The 5' ESTs may be to obtain cDNAs and genomic DNAs corresponding to the 5' ESTs. The 5' ESTs may also be used in diagnostic, forensic, gene therapy, and chromosome mapping procedures. Upstream regulatory sequences may also be obtained using the 5' ESTs. The 5' ESTs may also be used to design expression vectors and secretion vectors.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		



## 5' ESTS AND ENCODED HUMAN PROTEINS

Background of the Invention

The estimated 50,000-100,000 genes scattered along the human chromosomes offer tremendous  
5 promise for the understanding, diagnosis, and treatment of human diseases. In addition, probes capable  
of specifically hybridizing to loci distributed throughout the human genome find applications in the  
construction of high resolution chromosome maps and in the identification of individuals.

In the past, the characterization of even a single human gene was a painstaking process,  
requiring years of effort. Recent developments in the areas of cloning vectors, DNA sequencing, and  
10 computer technology have merged to greatly accelerate the rate at which human genes can be isolated,  
sequenced, mapped, and characterized.

Currently, two different approaches are being pursued for identifying and characterizing the  
genes distributed along the human genome. In one approach, large fragments of genomic DNA are  
isolated, cloned, and sequenced. Potential open reading frames in these genomic sequences are  
15 identified using bioinformatics software. However, this approach entails sequencing large stretches  
of human DNA which do not encode proteins in order to find the protein encoding sequences  
scattered throughout the genome. In addition to requiring extensive sequencing, the bioinformatics  
software may mischaracterize the genomic sequences obtained, *i.e.*, labeling non-coding DNA as  
coding DNA and vice versa.

20 An alternative approach takes a more direct route to identifying and characterizing human  
genes. In this approach, complementary DNAs (cDNAs) are synthesized from isolated messenger  
RNAs (mRNAs) which encode human proteins. Using this approach, sequencing is only performed on  
DNA which is derived from protein coding portions of the genome. Often, only short stretches of the  
cDNAs are sequenced to obtain sequences called expressed sequence tags (ESTs). The ESTs may then  
25 be used to isolate or purify extended cDNAs which include sequences adjacent to the EST sequences.  
The extended cDNAs may contain all of the sequence of the EST which was used to obtain them or only  
a portion of the sequence of the EST which was used to obtain them. In addition, the extended cDNAs  
may contain the full coding sequence of the gene from which the EST was derived or, alternatively, the  
extended cDNAs may include portions of the coding sequence of the gene from which the EST was  
30 derived. It will be appreciated that there may be several extended cDNAs which include the EST  
sequence as a result of alternate splicing or the activity of alternative promoters. Alternatively, ESTs  
having partially overlapping sequences may be identified and contigs comprising the consensus  
sequences of the overlapping ESTs may be identified.

In the past, these short EST sequences were often obtained from oligo-dT primed cDNA  
35 libraries. Accordingly, they mainly corresponded to the 3' untranslated region of the mRNA. In part,  
the prevalence of EST sequences derived from the 3' end of the mRNA is a result of the fact that typical

techniques for obtaining cDNAs, are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs (Adams *et al.*, *Nature* 377:3-174, 1996, Hillier *et al.*, *Genome Res.* 6:807-828, 1996).

In addition, in those reported instances where longer cDNA sequences have been obtained, the reported sequences typically correspond to coding sequences and do not include the full 5' untranslated region (5'UTR) of the mRNA from which the cDNA is derived. Indeed, 5'UTRs have been shown to affect either the stability or translation of mRNAs. Thus, regulation of gene expression may be achieved through the use of alternative 5'UTRs as shown, for instance, for the translation of the tissue inhibitor of metalloprotease mRNA in mitogenically activated cells (Waterhouse *et al.*, *J Biol Chem.* 265:5585-9, 1990). Furthermore, modification of 5'UTR through mutation, insertion or translocation events may even be implied in pathogenesis. For instance, the fragile X syndrome, the most common cause of inherited mental retardation, is partly due to an insertion of multiple CGG trinucleotides in the 5'UTR of the fragile X mRNA resulting in the inhibition of protein synthesis via ribosome stalling (Feng *et al.*, *Science* 268:731-4, 1995). An aberrant mutation in regions of the 5'UTR known to inhibit translation of the proto-oncogene *c-myc* was shown to result in upregulation of *c-myc* protein levels in cells derived from patients with multiple myelomas (Willis *et al.*, *Curr Top Microbiol Immunol* 224:269-76, 1997). In addition, the use of oligo-dT primed cDNA libraries does not allow the isolation of complete 5'UTRs since such incomplete sequences obtained by this process may not include the first exon of the mRNA, particularly in situations where the first exon is short. Furthermore, they may not include some exons, often short ones, which are located upstream of splicing sites. Thus, there is a need to obtain sequences derived from the 5' ends of mRNAs.

While many sequences derived from human chromosomes have practical applications, approaches based on the identification and characterization of those chromosomal sequences which encode a protein product are particularly relevant to diagnostic and therapeutic uses. In some instances, the sequences used in such therapeutic or diagnostic techniques may be sequences which encode proteins which are secreted from the cell in which they are synthesized. Those sequences encoding secreted proteins as well as the secreted proteins themselves, are particularly valuable as potential therapeutic agents. Such proteins are often involved in cell to cell communication and may be responsible for producing a clinically relevant response in their target cells. In fact, several secretory proteins, including tissue plasminogen activator, G-CSF, GM-CSF, erythropoietin, human growth hormone, insulin, interferon- $\alpha$ , interferon- $\beta$ , interferon- $\gamma$ , and interleukin-2, are currently in clinical use. These proteins are used to treat a wide range of conditions, including acute myocardial infarction, acute ischemic stroke, anemia, diabetes, growth hormone deficiency, hepatitis, kidney carcinoma, chemotherapy-induced neutropenia and multiple sclerosis. For these reasons, extended cDNAs encoding secreted proteins or portions thereof represent a valuable source of therapeutic agents. Thus, there is a need for the identification and characterization of secreted proteins and the nucleic acids encoding them.

In addition to being therapeutically useful themselves, secretory proteins include short peptides, called signal peptides, at their amino termini which direct their secretion. These signal peptides are

encoded by the signal sequences located at the 5' ends of the coding sequences of genes encoding secreted proteins. These signal peptides can be used to direct the extracellular secretion of any protein to which they are operably linked. In addition, portions of the signal peptides called membrane-translocating sequences, may also be used to direct the intracellular import of a peptide or protein of interest. This may prove beneficial in gene therapy strategies in which it is desired to deliver a particular gene product to cells other than the cells in which it is produced. Signal sequences encoding signal peptides also find application in simplifying protein purification techniques. In such applications, the extracellular secretion of the desired protein greatly facilitates purification by reducing the number of undesired proteins from which the desired protein must be selected. Thus, there exists a need to identify and characterize the 5' portions of the genes for secretory proteins which encode signal peptides.

Sequences coding for non-secreted proteins may also find application as therapeutics or diagnostics. In particular, such sequences may be used to determine whether an individual is likely to express a detectable phenotype, such as a disease, as a consequence of a mutation in the coding sequence of a protein. In instances where the individual is at risk of suffering from a disease or other undesirable phenotype as a result of a mutation in such a coding sequence, the undesirable phenotype may be corrected by introducing a normal coding sequence using gene therapy. Alternatively, if the undesirable phenotype results from overexpression of the protein encoded by the coding sequence, expression of the protein may be reduced using antisense or triple helix based strategies.

The secreted or non-secreted human polypeptides encoded by the coding sequences may also be used as therapeutics by administering them directly to an individual having a condition, such as a disease, resulting from a mutation in the sequence encoding the polypeptide. In such an instance, the condition can be cured or ameliorated by administering the polypeptide to the individual.

In addition, the secreted or non-secreted human polypeptides or portions thereof may be used to generate antibodies useful in determining the tissue type or species of origin of a biological sample. The antibodies may also be used to determine the cellular localization of the secreted or non-secreted human polypeptides or the cellular localization of polypeptides which have been fused to the human polypeptides. In addition, the antibodies may also be used in immunoaffinity chromatography techniques to isolate, purify, or enrich the human polypeptide or a target polypeptide which has been fused to the human polypeptide.

Public information on the number of human genes for which the promoters and upstream regulatory regions have been identified and characterized is quite limited. In part, this may be due to the difficulty of isolating such regulatory sequences. Upstream regulatory sequences such as transcription factor binding sites are typically too short to be utilized as probes for isolating promoters from human genomic libraries. Recently, some approaches have been developed to isolate human promoters. One of them consists of making a CpG island library (Cross *et al.*, *Nature Genetics* 6: 236-244, 1994). The second consists of isolating human genomic DNA sequences containing SpeI binding sites by the use of SpeI binding protein. (Mortlock *et al.*, *Genome Res.* 6:327-335, 1996). Both of these approaches have

their limits due to a lack of specificity and of comprehensiveness. Thus, there exists a need to identify and systematically characterize the 5' portions of the genes.

The present 5' ESTs may be used to efficiently identify and isolate 5'UTRs and upstream regulatory regions which control the location, developmental stage, rate, and quantity of protein synthesis, as well as the stability of the mRNA. Once identified and characterized, these regulatory regions may be utilized in gene therapy or protein purification schemes to obtain the desired amount and locations of protein synthesis or to inhibit, reduce, or prevent the synthesis of undesirable gene products.

In addition, ESTs containing the 5' ends of protein genes may include sequences useful as probes for chromosome mapping and the identification of individuals. Thus, there is a need to identify and characterize the sequences upstream of the 5' coding sequences of genes.

#### Summary of the Invention

The present invention relates to purified, isolated, or enriched 5' ESTs which include sequences derived from the authentic 5' ends of their corresponding mRNAs. The term "corresponding mRNA" refers to the mRNA which was the template for the cDNA synthesis which produced the 5' EST. These sequences will be referred to hereinafter as "5' ESTs." The present invention also includes purified, isolated or enriched nucleic acids comprising contigs assembled by determining a consensus sequences from a plurality of ESTs containing overlapping sequences. These contigs will be referred to herein as "consensus contigated 5'ESTs."

As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual 5' EST clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. The cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately  $10^4$ - $10^6$  fold purification of the native message. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

As used herein, the term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide present in a living animal is not isolated, but the same polynucleotide, separated from some or all of the coexisting materials in the natural system, is isolated.

As used herein, the term "recombinant" means that the 5' EST is adjacent to "backbone" nucleic acid to which it is not adjacent in its natural environment. Additionally, to be "enriched" the 5' ESTs will

represent 5% or more of the number of nucleic acid inserts in a population of nucleic acid backbone molecules. Backbone molecules according to the present invention include nucleic acids such as expression vectors, self-replicating nucleic acids, viruses, integrating nucleic acids, and other vectors or nucleic acids used to maintain or manipulate a nucleic acid insert of interest. Preferably, the enriched 5' ESTs represent 15% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. More preferably, the enriched 5' ESTs represent 50% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. In a highly preferred embodiment, the enriched 5' ESTs represent 90% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules.

10 "Stringent," "moderate," and "low" hybridization conditions are as defined below.

The term "polypeptide" refers to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude post-expression modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

As used interchangeably herein, the terms "nucleic acids," "oligonucleotides," and "polynucleotides" include RNA, DNA, or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form. The term "nucleotide" as used herein as an adjective to describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in single-stranded or duplex form. The term "nucleotide" is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. Although the term "nucleotide" is also used herein to encompass "modified nucleotides" which comprise at least one modifications (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar, for examples of analogous linking groups, purine, pyrimidines, and sugars see for example PCT publication No. WO 95/04064. The polynucleotide sequences of the invention may be prepared by any known method, including synthetic, recombinant, *ex vivo* generation, or a combination thereof, as well as utilizing any purification methods known in the art.

The terms "base paired" and "Watson & Crick base paired" are used interchangeably herein to refer to nucleotides which can be hydrogen bonded to one another by virtue of their sequence

identities in a manner like that found in double-helical DNA with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds (See Stryer, L., *Biochemistry*, 4<sup>th</sup> edition, 1995).

The terms "complementary" or "complement thereof" are used herein to refer to the  
5 sequences of polynucleotides which are capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. For the purpose of the present invention, a first polynucleotide is deemed to be complementary to a second polynucleotide when each base in the first polynucleotide is paired with its complementary base. Complementary bases are, generally, A and T (or A and U), or C and G. "Complement" is used  
10 herein as a synonym from "complementary polynucleotide," "complementary nucleic acid" and "complementary nucleotide sequence". These terms are applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind. Preferably, a "complementary" sequence is a sequence which an A at each position where there is a T on the opposite strand, a T at each position where there is an A on  
15 the opposite strand, a G at each position where there is a C on the opposite strand and a C at each position where there is a G on the opposite strand.

Thus, 5' ESTs in cDNA libraries in which one or more 5' ESTs make up 5% or more of the number of nucleic acid inserts in the backbone molecules are "enriched recombinant 5' ESTs" as defined herein. Likewise, 5' ESTs in a population of plasmids in which one or more 5' ESTs of the present  
20 invention have been inserted such that they represent 5% or more of the number of inserts in the plasmid backbone are "enriched recombinant 5' ESTs" as defined herein. However, 5' ESTs in cDNA libraries in which 5' ESTs constitute less than 5% of the number of nucleic acid inserts in the population of backbone molecules, such as libraries in which backbone molecules having a 5' EST insert are extremely rare, are not "enriched recombinant 5' ESTs."

25 In some embodiments, the present invention relates to 5' ESTs which are derived from genes encoding secreted proteins. As used herein, a "secreted" protein is one which, when expressed in a suitable host cell, is transported across or through a membrane, including transport as a result of signal peptides in its amino acid sequence. "Secreted" proteins include without limitation proteins secreted wholly (e.g. soluble proteins), or partially (e.g. receptors) from the cell in which they are expressed.  
30 "Secreted" proteins also include without limitation proteins which are transported across the membrane of the endoplasmic reticulum.

Such 5' ESTs include nucleic acid sequences, called signal sequences, which encode signal peptides which direct the extracellular secretion of the proteins encoded by the genes from which the 5' ESTs are derived. Generally, the signal peptides are located at the amino termini of secreted proteins.

35 Secreted proteins are translated by ribosomes associated with the "rough" endoplasmic reticulum. Generally, secreted proteins are co-translationally transferred to the membrane of the endoplasmic reticulum. Association of the ribosome with the endoplasmic reticulum during translation

of secreted proteins is mediated by the signal peptide. The signal peptide is typically cleaved following its co-translational entry into the endoplasmic reticulum. After delivery to the endoplasmic reticulum, secreted proteins may proceed through the Golgi apparatus. In the Golgi apparatus, the proteins may undergo post-translational modification before entering secretory vesicles which transport them across  
5 the cell membrane.

The 5' ESTs of the present invention have several important applications. For example, they may be used to obtain and express cDNA clones which include the full protein coding sequences of the corresponding gene products, including the authentic translation start sites derived from the 5' ends of the coding sequences of the mRNAs from which the 5' ESTs are derived. These cDNAs will be referred  
10 to hereinafter as "full-length cDNAs." These cDNAs may also include DNA derived from mRNA sequences upstream of the translation start site. The full-length cDNA sequences may be used to express the proteins corresponding to the 5' ESTs. As discussed above, secreted proteins and non-secreted proteins may be therapeutically important. Thus, the proteins expressed from the cDNAs may be useful in treating and controlling a variety of human conditions. The 5' ESTs may also be used to obtain the  
15 corresponding genomic DNA. The term "corresponding genomic DNA" refers to the genomic DNA which encodes the mRNA from which the 5' EST was derived.

Alternatively, the 5' ESTs may be used to obtain and express extended cDNAs encoding portions of the protein. In the case of secreted proteins, the portions may comprise the signal peptides of the secreted proteins or the mature proteins generated when the signal peptide is cleaved off.  
20 The present invention includes isolated, purified, or enriched "EST-related nucleic acids." The terms "isolated," "purified" or "enriched" have the meanings provided above. As used herein, the term "EST-related nucleic acids" means the nucleic acids of SEQ ID NOs. 24-811 and 1600-1622, extended cDNAs obtainable using the nucleic acids of SEQ ID NOs. 24-811 and 1600-1622, full-length cDNAs obtainable using the nucleic acids of SEQ ID NOs. 24-811 and 1600-1622 or genomic DNAs obtainable  
25 using the nucleic acids of SEQ ID NOs. 24-811 and 1600-1622. The present invention also includes the sequences complementary to the EST-related nucleic acids.

The present invention also includes isolated, purified, or enriched "fragments of EST-related nucleic acids." The terms "isolated," "purified" and "enriched" have the meanings described above. As used herein the term "fragments of EST-related nucleic acids" means fragments comprising at least 10,  
30 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, 50, 75, 100, 200, 300, 500, or 1000 consecutive nucleotides of the EST-related nucleic acids to the extent that fragments of these lengths are consistent with the lengths of the particular EST-related nucleic acids being referenced. In particular, fragments of EST-related nucleic acids refer to "polynucleotides described in Table II," "polynucleotides described in Table III," and "polynucleotides described in Table IV." The present invention also includes the sequences  
35 complementary to the fragments of the EST-related nucleic acids.

The present invention also includes isolated, purified, or enriched "positional segments of EST-related nucleic acids." As used herein, the term "positional segments of EST-related nucleic acids"

includes segments comprising nucleotides 1-25, 26-50, 51-75, 76-100, 101-125, 126-150, 151-175, 176-200, 201-225, 226-250, 251-300, 301-325, 326-350, 351-375, 376-400, 401-425, 426-450, 451-475, 476-500, 501-525, 526-550, 551-575, 576-600 and 601-the terminal nucleotide of the EST-related nucleic acids to the extent that such nucleotide positions are consistent with the lengths of the particular

5 EST-related nucleic acids being referenced. The term "positional segments of EST-related nucleic acids" also includes segments comprising nucleotides 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 450-500, 501-550, 551-600 or 601-the terminal nucleotide of the EST-related nucleic acids to the extent that such nucleotide positions are consistent with the lengths of the particular EST-related nucleic acids being referenced. The term "positional segments of EST-related nucleic

10 acids" also includes segments comprising nucleotides 1-100, 101-200, 201-300, 301-400, 501-500, 500-600, or 601-the terminal nucleotide of the EST-related nucleic acids to the extent that such nucleotide positions are consistent with the lengths of the particular EST-related nucleic acids being referenced. In addition, the term "positional segments of EST-related nucleic acids" includes segments comprising nucleotides 1-200, 201-400, 400-600, or 601-the terminal nucleotide of the EST-related nucleic acids to

15 the extent that such nucleotide positions are consistent with the lengths of the particular EST-related nucleic acids being referenced. The present invention also includes the sequences complementary to the positional segments of EST-related nucleic acids.

The present invention also includes isolated, purified, or enriched "fragments of positional segments of EST-related nucleic acids." As used herein, the term "fragments of positional segments of

20 EST-related nucleic acids" refers to fragments comprising at least 10, 15, 18, 20, 23, 25, 28, 30, 35, 40, 50, 75, 100, 150, or 200 consecutive nucleotides of the positional segments of EST-related nucleic acids. The present invention also includes the sequences complementary to the fragments of positional segments of EST-related nucleic acids.

The present invention also includes isolated or purified "EST-related polypeptides." As used

25 herein, the term "EST-related polypeptides" means the polypeptides encoded by the EST-related nucleic acids, including the polypeptides of SEQ ID NOs. 812-1599.

The present invention also includes isolated or purified "fragments of EST-related polypeptides." As used herein, the term "fragments of EST-related polypeptides" means fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids of an EST-

30 related polypeptide to the extent that fragments of these lengths are consistent with the lengths of the particular EST-related polypeptides being referenced. In particular, fragments of EST-related polypeptides refer to polypeptides encoded by "polynucleotides described in Table II," "polynucleotides described in Table III," and "polynucleotides described in Table IV."

The present invention also includes isolated or purified "positional segments of EST-related

35 polypeptides." As used herein, the term "positional segments of EST-related polypeptides" includes polypeptides comprising amino acid residues 1-25, 26-50, 51-75, 76-100, 101-125, 126-150, 151-175, 176-200, or 201-the C-terminal amino acid of the EST-related polypeptides to the extent that such amino



acid residues are consistent with the lengths of the particular EST-related polypeptides being referenced. The term "positional segments of EST-related polypeptides" also includes segments comprising amino acid residues 1-50, 51-100, 101-150, 151-200 or 201-the C-terminal amino acid of the EST-related polypeptides to the extent that such amino acid residues are consistent with the lengths of the particular

5 EST-related polypeptides being referenced. The term "positional segments of EST-related polypeptides" also includes segments comprising amino acids 1-100 or 101-200 of the EST-related polypeptides to the extent that such amino acid residues are consistent with the lengths of particular EST-related polypeptides being referenced. In addition, the term "positional segments of EST-related polypeptides" includes segments comprising amino acid residues 1-200 or 201-the C-terminal amino acid of the EST-

10 related polypeptides to the extent that amino acid residues are consistent with the lengths of the particular EST-related polypeptides being referenced.

The present invention also includes isolated or purified "fragments of positional segments of EST-related polypeptides." As used herein, the term "fragments of positional segments of EST-related polypeptides" means fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150

15 consecutive amino acids of positional segments of EST-related polypeptides to the extent that fragments of these lengths are consistent with the lengths of the particular EST-related polypeptides being referenced.

The present invention also includes antibodies which specifically recognize the EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides,

20 or fragments of positional segments of EST-related polypeptides. In the case of secreted proteins, such as those of SEQ ID NOs. 1554-1580 antibodies which specifically recognize the mature protein generated when the signal peptide is cleaved may also be obtained as described below. Similarly, antibodies which specifically recognize the signal peptides of SEQ ID NOs. 812-1516 or 1554-1580 may also be obtained.

25 In some embodiments and in the case of secreted proteins, the EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids, or fragments of positional segments of nucleic acids include a signal sequence. In other embodiments, the EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids, or fragments of positional segments of nucleic acids may include the full coding sequence for the

30 protein or, in the case of secreted proteins, the full coding sequence of the mature protein (*i.e.* the protein generated when the signal polypeptide is cleaved off). In addition, the EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids, or fragments of positional segments of nucleic acids may include regulatory regions upstream of the translation start site or downstream of the stop codon which control the amount, location, or developmental stage of gene

35 expression.

As discussed above, both secreted and non-secreted human proteins may be therapeutically important. Thus, the proteins expressed from the EST-related nucleic acids, fragments of EST-related

nucleic acids, positional segments of EST-related nucleic acids, or fragments of positional segments of nucleic acids may be useful in treating or controlling a variety of human conditions.

The EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids, or fragments of positional segments of nucleic acids may be used in forensic  
5 procedures to identify individuals or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal gene expression. In addition, the EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids, or fragments of positional segments of nucleic acids are useful for constructing a high resolution map of the human chromosomes.

10 The present invention also relates to secretion vectors capable of directing the secretion of a protein of interest. Such vectors may be used in gene therapy strategies in which it is desired to produce a gene product in one cell which is to be delivered to another location in the body. Secretion vectors may also facilitate the purification of desired proteins.

The present invention also relates to expression vectors capable of directing the expression of an  
15 inserted gene in a desired spatial or temporal manner or at a desired level. Such vectors may include sequences upstream of the EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids, or fragments of positional segments of nucleic acids, such as promoters or upstream regulatory sequences.

The present invention also comprises fusion vectors for making chimeric polypeptides  
20 comprising a first polypeptide and a second polypeptide. Such vectors are useful for determining the cellular localization of the chimeric polypeptides or for isolating, purifying or enriching the chimeric polypeptides.

The EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids, or fragments of positional segments of nucleic acids may also be used for  
25 gene therapy to control or treat genetic diseases. In the case of secreted proteins, signal peptides may be fused to heterologous proteins to direct their extracellular secretion.

Bacterial clones containing Bluescript plasmids having inserts containing the sequence of the non-aligned 5'ESTs, also referred to as singletons, and sequences of the 5'ESTs which were aligned to yield consensus contigated 5' ESTs are presently stored at 80°C in 4% (v/v) glycerol in the inventor's  
30 laboratories under internal designations. The non-aligned 5'ESTs are those which comprise a single EST from a single tissue in the listing of Table V. The inserts may be recovered from the stored materials by growing the appropriate clones on a suitable medium. The Bluescript DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be  
35 further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR

can be performed with primers designed at both ends of the inserted EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids, or fragments of positional segments of nucleic acids. The PCR product which corresponds to the EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids, or fragments of positional segments of nucleic acids can then be manipulated using standard cloning techniques familiar to those skilled in the art.

One embodiment of the present invention is a purified nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and sequences complementary to the sequences of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622.

Another embodiment of the present invention is a purified nucleic acid comprising at least 10, 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, 50, 75, 100, 200, 300, 500, or 1000 consecutive nucleotides, to the extent that fragments of these lengths are consistent with the specific sequence, of a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and sequences complementary to the sequences of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622.

A further embodiment of the present invention is a purified nucleic acid comprising the coding sequence of a sequence selected from the group consisting of SEQ ID NOs. 24-811.

Yet another embodiment of the present invention is a purified nucleic acid comprising the full coding sequences of a sequence selected from the group consisting of SEQ ID NOs. 766-792 wherein the full coding sequence comprises the sequence encoding the signal peptide and the sequence encoding the mature protein.

Still another embodiment of the present invention is a purified nucleic acid comprising a contiguous span of a sequence selected from the group consisting of SEQ ID NOs. 766-792 which encodes the mature protein.

Another embodiment of the present invention is a purified nucleic acid comprising a contiguous span of a sequence selected from the group consisting of SEQ ID NOs. 24-728 and 766-792 which encodes the signal peptide.

Another embodiment of the present invention is a purified nucleic acid encoding a polypeptide comprising a sequence selected from the group consisting of the sequences of SEQ ID NOs. 812-1599.

Another embodiment of the present invention is a purified nucleic acid encoding a polypeptide comprising a sequence selected from the group consisting of the sequences of SEQ ID NOs. 1554-1580.

Another embodiment of the present invention is a purified nucleic acid encoding a polypeptide comprising a mature protein included in a sequence selected from the group consisting of the sequences of SEQ ID NOs. 1554-1580.

Another embodiment of the present invention is a purified nucleic acid encoding a polypeptide comprising a signal peptide included in a sequence selected from the group consisting of the sequences of SEQ ID NOs. 812-1516 and 1554-1580.

Another embodiment of the present invention is a purified nucleic acid at least 30, 35, 40, 50,  
5 75, 100, 200, 300, 500 or 1000 nucleotides in length which hybridizes under stringent conditions to a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and sequences complementary to the sequences of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622.

Another embodiment of the present invention is a purified or isolated polypeptide comprising  
10 a sequence selected from the group consisting of the sequences of SEQ ID NOs. 812-1599.

Another embodiment of the present invention is a purified or isolated polypeptide comprising a sequence selected from the group consisting of SEQ ID NOs. 1554-1580.

Another embodiment of the present invention is a purified or isolated polypeptide comprising a mature protein of a polypeptide selected from the group consisting of SEQ ID NOs. 1554-1580.

15 Another embodiment of the present invention is a purified or isolated polypeptide comprising a signal peptide of a sequence selected from the group consisting of the polypeptides of SEQ ID NOs. 812-1516 and 1554-1580.

Another embodiment of the present invention is a purified or isolated polypeptide comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, 50, 75, 100, 200, 300, 500, or 1000 consecutive amino  
20 acids, to the extent that fragments of these lengths are consistent with the specific sequence, of a sequence selected from the group consisting of the sequences of SEQ ID NOs. 812-1599.

Another embodiment of the present invention is a method of making a cDNA comprising the steps of contacting a collection of mRNA molecules from human cells with a primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence selected from  
25 the group consisting of the sequences complementary to SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622, hybridizing said primer to an mRNA in said collection that encodes said protein reverse transcribing said hybridized primer to make a first cDNA strand from said mRNA, making a second cDNA strand complementary to said first cDNA strand and isolating the resulting cDNA encoding said protein comprising said first cDNA strand and said second cDNA strand.

30 Another embodiment of the present invention is a purified cDNA obtainable by the method of the preceding paragraph.

In one aspect of this embodiment, the cDNA encodes at least a portion of a human polypeptide.

Another embodiment of the present invention is a method of making a cDNA comprising the  
35 steps of obtaining a cDNA comprising a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622, contacting said cDNA with a detectable probe comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence

selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and the sequences complementary to SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 under conditions which permit said probe to hybridize to said cDNA, identifying a cDNA which hybridizes to said detectable probe, and isolating said cDNA which hybridizes to said probe.

5 Another embodiment of the present invention is a purified cDNA obtainable by the method of the preceding paragraph.

In one aspect of this embodiment, the cDNA encodes at least a portion of a human polypeptide.

Another embodiment of the present invention is a method of making a cDNA comprising the  
10 steps of contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA, hybridizing said first primer to said polyA tail, reverse transcribing said mRNA to make a first cDNA strand, making a second cDNA strand complementary to said first cDNA strand using at least one primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID  
15 NOs. 24-811 and SEQ ID NOs. 1600-1622, and isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

Another embodiment of the present invention is a purified cDNA obtainable by the method of the preceding paragraph.

In one aspect of this embodiment, said cDNA encodes at least a portion of a human  
20 polypeptide.

In another aspect of the preceding method the second cDNA strand is made by contacting said first cDNA strand with a first pair of primers, said first pair of primers comprising a second primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622  
25 and a third primer having a sequence therein which is included within the sequence of said first primer, performing a first polymerase chain reaction with said first pair of primers to generate a first PCR product, contacting said first PCR product with a second pair of primers, said second pair of primers comprising a fourth primer, said fourth primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of said sequence selected from the group consisting of SEQ  
30 ID NOs. 24-811 and SEQ ID NOs. 1600-1622, and a fifth primer, wherein said fourth and fifth hybridize to sequences within said first PCR product, and performing a second polymerase chain reaction, thereby generating a second PCR product.

One aspect of this embodiment is a purified cDNA obtainable by the method of the preceding paragraph.

35 In another aspect of this embodiment, said cDNA encodes at least a portion of a human polypeptide.

Alternatively, the second cDNA strand may be made by contacting said first cDNA strand with a second primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622, hybridizing said second primer to said first strand cDNA, and extending said  
5 hybridized second primer to generate said second cDNA strand.

One aspect of the above embodiment is a purified cDNA obtainable by the method of the preceding paragraph.

In a further aspect of this embodiment said cDNA encodes at least a portion of a human polypeptide.

10 Another embodiment of the present invention is a method of making a polypeptide comprising the steps of obtaining a cDNA which encodes a polypeptide encoded by a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs. 24-811 or a cDNA which encodes a polypeptide comprising at least 6, 8, 10, 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive amino acids of a polypeptide encoded by a sequence selected from the group consisting  
15 of SEQ ID NOs. 24-811, inserting said cDNA in an expression vector such that said cDNA is operably linked to a promoter, introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said cDNA, and isolating said protein.

Another aspect of this embodiment is an isolated protein obtainable by the method of the preceding paragraph.

20 Another embodiment of the present invention is a method of obtaining a promoter DNA comprising the steps of obtaining genomic DNA located upstream of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and the sequences complementary to the sequences of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622, screening said genomic DNA to identify a promoter capable of directing transcription  
25 initiation, and isolating said DNA comprising said identified promoter.

In one aspect of this embodiment, said obtaining step comprises walking from genomic DNA comprising a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and the sequences complementary to SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622. In another aspect of this embodiment, said screening step comprises inserting genomic DNA located  
30 upstream of a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and the sequences complementary to SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 into a promoter reporter vector. For example, said screening step may comprise identifying motifs in genomic DNA located upstream of a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and the sequences complementary to SEQ ID NOs.  
35 24-811 and SEQ ID NOs. 1600-1622 which are transcription factor binding sites or transcription start sites.

Another embodiment of the present invention is a isolated promoter obtainable by the method of the paragraph above.

Another embodiment of the present invention is the inclusion of at least one sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622, the  
5 sequences complementary to the sequences of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and fragments comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, 50, or 100 consecutive nucleotides of said sequence in an array of discrete ESTs or fragments thereof of at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, 50, or 100 nucleotides in length. In some aspects of this embodiment, the array includes at least two sequences selected from the group consisting of SEQ ID NOs. 24-811 and  
10 SEQ ID NOs. 1600-1622, the sequences complementary to the sequences of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622, and fragments comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, 50, or 100 consecutive nucleotides of said sequences. In another aspect of this embodiment., the array includes at least one, three, five, ten, fifteen, or twenty sequences selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622, the sequences complementary to  
15 the sequences of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and fragments comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, 50, or 100 consecutive nucleotides of said sequences.

Another embodiment of the present invention is an enriched population of recombinant nucleic acids, said recombinant nucleic acids comprising an insert nucleic acid and a backbone nucleic acid, wherein at least 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 5%, 10%, or 20% of said insert  
20 nucleic acids in said population comprise a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and the sequences complementary to SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622.

Another embodiment of the present invention is a purified or isolated antibody capable of specifically binding to a polypeptide comprising a sequence selected from the group consisting of  
25 SEQ ID NOs. 812-1599.

Another embodiment of the present invention is a purified or isolated antibody capable of specifically binding to a polypeptide comprising at least 6, 8, 10, 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive amino acids of a sequence selected from the group consisting of SEQ ID NOs. 812-1599.

30 Yet, another embodiment of the present invention is an antibody composition capable of selectively binding to an epitope-containing fragment of a polypeptide comprising a contiguous span of at least 8, 10, 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 amino acids of any of SEQ ID NOs. 812-1599, wherein said antibody is polyclonal or monoclonal.

Another embodiment of the present invention is a computer readable medium having stored  
35 thereon a sequence selected from the group consisting of a nucleic acid code of SEQ ID NOs. 24-811 and 1600-1622 and a polypeptide code of SEQ ID NOs. 812-1599.

Another embodiment of the present invention is a computer system comprising a processor and a data storage device wherein said data storage device has stored thereon a sequence selected from the group consisting of a nucleic acid code of SEQID NOs. 24-811 and 1600-1622 and a polypeptide code of SEQ ID NOs. 812-1599. In one aspect of this embodiment the computer system  
5 further comprises a sequence comparer and a data storage device having reference sequences stored thereon. For example, the sequence comparer may comprise a computer program which indicates polymorphisms. In another aspect of this embodiment, the computer system further comprises an identifier which identifies features in said sequence.

Another embodiment of the present invention is a method for comparing a first sequence to a  
10 reference sequence wherein said first sequence is selected from the group consisting of a nucleic acid code of SEQID NOs. 24-811 and 1600-1622 and a polypeptide code of SEQ ID NOs. 812-1599 comprising the steps of reading said first sequence and said reference sequence through use of a computer program which compares sequences and determining differences between said first sequence and said reference sequence with said computer program. In some aspects of this embodiment, said step  
15 of determining differences between the first sequence and the reference sequence comprises identifying polymorphisms.

Another embodiment of the present invention is a method for identifying a feature in a sequence selected from the group consisting of a nucleic acid code of SEQID NOs. 24-811 and 1600-1622 and a polypeptide code of SEQ ID NOs. 812-1599 comprising the steps of reading said  
20 sequence through the use of a computer program which identifies features in sequences and identifying features in said sequence with said computer program.

Another embodiment of the present invention is a vector comprising a nucleic acid according to any one of the nucleic acids described above.

Another embodiment of the present invention is a host cell containing the above vector.

25 Another embodiment of the present invention is a method of making any of the nucleic acids described above comprising the steps of introducing said nucleic acid into a host cell such that said nucleic acid is present in multiple copies in each host cell and isolating said nucleic acid from said host cell.

Another embodiment of the present invention is a method of making a nucleic acid of any of  
30 the nucleic acids described above comprising the step of sequentially linking together the nucleotides in said nucleic acids.

Another embodiment of the present invention is a method of making any of the polypeptides described above wherein said polypeptides is 150 amino acids in length or less comprising the step of sequentially linking together the amino acids in said polypeptide.

35 Another embodiment of the present invention is a method of making any of the polypeptides described above wherein said polypeptides is 120 amino acids in length or less comprising the step of sequentially linking together the amino acids in said polypeptides.



### Brief Description of the Drawings

Figure 1 is a summary of a procedure for obtaining cDNAs which have been selected to include the 5' ends of the mRNAs from which they derived. In the first step (1), the cap of intact mRNAs is oxidized to be chemically ligated to an oligonucleotide tag. In the second step (2), a reverse transcription is performed using random primers to generate a first cDNA strand. In the third step (3), mRNAs are eliminated and the second strand synthesis is carried out using a primer contained in the oligonucleotide tag.

Figure 2 is an analysis of the 43 amino terminal amino acids of all human SwissProt proteins to determine the frequency of false positives and false negatives using the techniques for signal peptide identification described herein.

Figure 3 summarizes a general method used to clone and sequence extended cDNAs containing sequences adjacent to 5'ESTs.

Figure 4 provides a schematic description of the promoters isolated and the way they are assembled with the corresponding 5' tags.

Figure 5 describes the transcription factor binding sites present in each of the promoters of Figure 4.

Figure 6 is a block diagram of an exemplary computer system.

Figure 7 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database.

Figure 8 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous.

Figure 9 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence.

Figure 10 is a table with all of the parameters that can be used for each step of extended cDNA analysis.

### Detailed Description of the Preferred Embodiment

#### **I. Obtaining 5'ESTs from cDNA libraries including the 5'Ends of their Corresponding mRNAs**

The 5' ESTs of the present invention were obtained from cDNA libraries including cDNAs which include the 5'end of their corresponding mRNAs. The general method used to obtain such cDNA libraries is described in Examples 1 to 5.

#### **EXAMPLE 1**

##### Preparation of mRNA

Total human RNAs or polyA<sup>+</sup> RNAs derived from 29 different tissues were respectively purchased from LABIMO and CLONTECH and used to generate 44 cDNA libraries as described below.

The purchased RNA had been isolated from cells or tissues using acid guanidium thiocyanate-phenol-chloroform extraction (Chomczynski and Sacchi, *Analytical Biochemistry* 162:156-159, 1987). PolyA<sup>+</sup> RNA was isolated from total RNA (LABIMO) by two passes of oligo dT chromatography, as described by Aviv and Leder, *Proc. Natl. Acad. Sci. USA* 69:1408-1412, 1972) in order to eliminate ribosomal  
5 RNA.

The quality and the integrity of the polyA<sup>+</sup> RNAs were checked. Northern blots hybridized with a globin probe were used to confirm that the mRNAs were not degraded. Contamination of the polyA<sup>+</sup> mRNAs by ribosomal sequences was checked using Northern blots and a probe derived from the sequence of the 28S rRNA. Preparations of mRNAs with less than 5% of rRNAs were used in  
10 library construction. To avoid constructing libraries with RNAs contaminated by exogenous sequences (prokaryotic or fungal), the presence of bacterial 16S ribosomal sequences or of two highly expressed fungal mRNAs was examined using PCR.

## EXAMPLE 2

### 15 Methods for Obtaining mRNAs having Intact 5' Ends

Following preparation of the mRNAs from various tissues as described above, selection of mRNA with intact 5' ends and specific attachment of an oligonucleotide tag to the 5' end of such mRNA was performed using either a chemical or enzymatic approach. Both techniques takes advantage of the presence of the "cap" structure, which characterizes the 5' end of intact mRNAs and which comprises a  
20 guanosine generally methylated once, at the 7 position. The chemical approach is illustrated in Figure 1.

The chemical modification approach involves the optional elimination of the 2', 3'-cis diol of the 3' terminal ribose, the oxidation of the 2', 3', -cis diol of the ribose linked to the cap of the 5' ends of the mRNAs into a dialdehyde, and the coupling of the such obtained dialdehyde to a derivatized oligonucleotide tag. Further detail regarding the chemical approaches for obtaining mRNAs having  
25 intact 5' ends are disclosed in International Application No. WO96/34981, published November 7, 1996.

The enzymatic approach for ligating the oligonucleotide tag to the 5' ends of mRNAs with intact 5' ends involves the removal of the phosphate groups present on the 5' ends of uncapped incomplete mRNAs, the subsequent decapping of mRNAs with intact 5' ends and the ligation of the phosphate present at the 5' end of the decapped mRNA to an oligonucleotide tag. Further detail regarding the  
30 enzymatic approaches for obtaining mRNAs having intact 5' ends are disclosed in Dumas Milne Edwards J.B. (Doctoral Thesis of Paris VI University, Le clonage des ADNc complets: difficultes et perspectives nouvelles. Apports pour l'etude de la regulation de l'expression de la tryptophane hydroxylase de rat, 20 Dec. 1993), EP0 625572 and Kato *et al.*, *Gene* 150:243-250 (1994).

In either the chemical or the enzymatic approach, the oligonucleotide tag has a restriction  
35 enzyme site (e.g. EcoRI sites) therein to facilitate later cloning procedures. Following attachment of the oligonucleotide tag to the mRNA, the integrity of the mRNA was then examined by performing a Northern blot using a probe complementary to the oligonucleotide tag.

**EXAMPLE 3****cDNA Synthesis Using mRNA Templates Having Intact 5' Ends**

For the mRNAs joined to oligonucleotide tags, first strand cDNA synthesis was performed using  
5 a reverse transcriptase with random nonamers as primers. In order to protect internal EcoRI sites in the  
cDNA from digestion at later steps in the procedure, methylated dCTP was used for first strand  
synthesis. After removal of mRNA by an alkaline hydrolysis, the first strand of cDNA was precipitated  
using isopropanol in order to eliminate residual primers.

The second strand of the cDNA was synthesized with a Klenow fragment using a primer  
10 corresponding to the 5' end of the ligated oligonucleotide. Methylated dCTP was also used for second  
strand synthesis in order to protect internal EcoRI sites in the cDNA from digestion during the cloning  
process.

**EXAMPLE 4****Cloning of cDNAs derived from mRNA with intact 5' ends into BlueScript**

Following second strand synthesis, the ends of the cDNA were blunted with T4 DNA  
polymerase (Biolabs) and the cDNA was digested with EcoRI. Since methylated dCTP was used during  
cDNA synthesis, the EcoRI site present in the tag was the only hemi-methylated site, hence the only site  
susceptible to EcoRI digestion. The cDNA was then size fractionated using exclusion chromatography  
20 (AcA, Biosepra) and fractions corresponding to cDNAs of more than 150 bp were pooled and ethanol  
precipitated. The cDNA was directionally cloned into the SmaI and EcoRI ends of the phagemid  
pBlueScript vector (Stratagene). The ligation mixture was electroporated into bacteria and propagated  
under appropriate antibiotic selection.

**EXAMPLE 5****Selection of Clones Having the Oligonucleotide Tag Attached Thereto**

Clones containing the oligonucleotide tag attached were then selected as follows. The plasmid  
DNAs containing 5' EST libraries made as described above were purified (Qiagen). A positive selection  
of the tagged clones was performed as follows. Briefly, in this selection procedure, the plasmid DNA  
30 was converted to single stranded DNA using gene II endonuclease of the phage F1 in combination with  
an exonuclease (Chang *et al.*, **Gene** 127:95-8, 1993) such as exonuclease III or T7 gene 6 exonuclease.  
The resulting single stranded DNA was then purified using paramagnetic beads as described by Fry *et al.*,  
*Biotechniques*, 13: 124-131, 1992. In this procedure, the single stranded DNA was hybridized with a  
biotinylated oligonucleotide having a sequence corresponding to the 3' end of the oligonucleotide tag.  
35 Clones including a sequence complementary to the biotinylated oligonucleotide were captured by  
incubation with streptavidin coated magnetic beads followed by magnetic selection. After capture of the  
positive clones, the plasmid DNA was released from the magnetic beads and converted into double

stranded DNA using a DNA polymerase such as the ThermoSequenase obtained from Amersham Pharmacia Biotech. The double stranded DNA was then electroporated into bacteria. The percentage of positive clones having the 5' tag oligonucleotide was estimated using dot blot analysis to typically be between 90 and 98%.

5       Following electroporation, the libraries were ordered in 384-microtiter plates (MTP). A copy of the MTP was stored for future needs. Then the libraries were transferred into 96 MTP and sequenced as described below.

#### EXAMPLE 6

##### 10       Sequencing of Inserts in Selected Clones

Plasmid inserts were first amplified by PCR on PE-9600 thermocyclers (Perkin-Elmer, Applied Biosystems Division, Foster City, CA), using standard SETA-A and SETA-B primers (Genset SA), AmpliTaqGold (Perkin-Elmer), dNTPs (Boehringer), buffer and cycling conditions as recommended by the Perkin-Elmer Corporation.

15       PCR products were then sequenced using automatic ABI Prism 377 sequencers (Perkin Elmer). Sequencing reactions were performed using PE 9600 thermocyclers with standard dye-primer chemistry and ThermoSequenase (Amersham Pharmacia Biotech). The primers used were either T7 or 21M13 (available from Genset SA) as appropriate. The primers were labeled with the JOE, FAM, ROX and TAMRA dyes. The dNTPs and ddNTPs used in the sequencing reactions were purchased from  
20   Boehringer. Sequencing buffer, reagent concentrations and cycling conditions were as recommended by Amersham.

Following the sequencing reaction, the samples were precipitated with ethanol, resuspended in formamide loading buffer, and loaded on a standard 4% acrylamide gel. Electrophoresis was performed for 2.5 hours at 3000V on an ABI 377 sequencer, and the sequence data were collected and analyzed  
25   using the ABI Prism DNA Sequencing Analysis Software, version 2.1.2.

#### EXAMPLE 7

##### Obtaining 5' ESTs from Extended cDNA libraries

##### Obtained from mRNA with Intact 5' Ends

30       Alternatively, 5'ESTs may be isolated from other cDNA or genomic DNA libraries. Such cDNA or genomic DNA libraries may be obtained from a commercial source or made using other techniques familiar to those skilled in the art. One example of such cDNA library construction, a full-length cDNA library, is as follows.

PolyA+ RNAs are prepared and their quality checked as described in Example 1. Then, the  
35   caps at the 5' ends of the polyA+ RNAs are specifically joined to an oligonucleotide tag as described in Example 2. The oligonucleotide tag may contain a restriction site such as Eco RI to facilitate further

subcloning procedures. Northern blotting is then performed to check the size of mRNAs having the oligonucleotide tag attached thereto and to ensure that the mRNAs are actually tagged.

First strand synthesis is subsequently carried out for mRNAs joined to the oligonucleotide tag as described in Example 3 above except that the random nonamers are replaced by an oligo-dT primer. For instance, this oligo-dT primer may contain an internal tag of 4 nucleotides which is different from one tissue to the other. Following second strand synthesis using a primer contained in the oligonucleotide tag attached to the 5' end of mRNA, the blunt ends of the obtained double stranded full-length DNAs are modified into cohesive ends to facilitate subcloning. For example, the extremities of full-length cDNAs may be modified to allow subcloning into the Eco RI and Hind III sites of a Bluescript vector using the Eco RI site of the oligonucleotide tag and the addition of a Hind III adaptor to the 3' end of full-length cDNAs.

The full-length cDNAs are then separated into several fractions according to their sizes using techniques familiar to those skilled in the art. For example, electrophoretic separation may be applied in order to yield 3 or 6 different fractions. Following gel extraction and purification, the cDNA fractions are subcloned into appropriate vectors, such as Bluescript vectors, transformed into competent bacteria and propagated under appropriate antibiotic conditions. Subsequently, plasmids containing tagged full-length cDNAs are positively selected as described in Example 5.

The 5' end of full-length cDNAs isolated from such cDNA libraries may then be sequenced as described in Example 6 to yield 5'ESTs.

20

## **II. Computer Analysis of the Isolated 5' ESTs: Construction of the SignalTag™ Database**

The sequence data from the cDNA libraries made as described above were transferred to a database, where quality control and validation steps were performed. A base-caller, working using a Unix system, automatically flagged suspect peaks, taking into account the shape of the peaks, the inter-peak resolution, and the noise level. The base-caller also performed an automatic trimming. Any stretch of 25 or fewer bases having more than 4 suspect peaks was considered unreliable and was discarded. Sequences corresponding to cloning vector or ligation oligonucleotides were automatically removed from the EST sequences. However, the resulting EST sequences may contain 1 to 5 bases belonging to the above mentioned sequences at their 5' end. If needed, these can easily be removed on a case to case basis.

Following sequencing as described above, the sequences of the 5' ESTs were entered in a database for storage and manipulation as described below. Before searching the ESTs in the database for sequences of interest, ESTs derived from mRNAs which were not of interest were identified. Briefly, such undesired sequences may be of three types. First, contaminants of either endogenous (ribosomal RNAs, transfer RNAs, mitochondrial RNAs) or exogenous (prokaryotic RNAs and fungal RNAs) origins were identified. Second, uninformative sequences, namely redundant sequences, small sequences and highly degenerate sequences were identified. Third, repeated sequences (Alu, L1, THE

35

and MER repeats, SSTR sequences or satellite, micro-satellite, or telomeric repeats) were identified and masked in further processing.

In order to determine the accuracy of the sequencing procedure as well as the efficiency of the 5' selection described above, the analyses described in Examples 8 and 9 respectively were performed on 5' ESTs obtained from the database following the elimination of endogenous and exogenous contaminants and following the masking of repeats.

### EXAMPLE 8

#### Measurement of Sequencing Accuracy by Comparison to Known Sequences

To further determine the accuracy of the sequencing procedure described in Example 6, the sequences of 5' ESTs derived from known sequences were identified and compared to the original known sequences. First, a FASTA analysis with overhangs shorter than 5 bp on both ends was conducted on the 5' ESTs to identify those matching an entry in the public human mRNA database available at the time of filing the priority applications. The 5' ESTs which matched a known human mRNA were then realigned with their cognate mRNA and dynamic programming was used to include substitutions, insertions, and deletions in the list of "errors" which would be recognized. Errors occurring in the last 10 bases of the 5' EST sequences were ignored to avoid the inclusion of spurious cloning sites in the analysis of sequencing accuracy. This analysis revealed that the sequences incorporated in the database had an accuracy of more than 99.5%.

20

### EXAMPLE 9

#### Determination of Efficiency of 5' EST Selection

To determine the efficiency at which the above selection procedures isolated 5' ESTs which included sequences close to the 5' end of the mRNAs from which they derived, the sequences of the ends of the 5' ESTs derived from the elongation factor 1 subunit  $\alpha$  and ferritin heavy chain genes were compared to the known cDNA sequences of these genes. Since the transcription start sites of both genes are well characterized, they may be used to determine the percentage of derived 5' ESTs which included the authentic transcription start sites. For both genes, more than 95% of the obtained 5' ESTs actually included sequences close to or upstream of the 5' end of the corresponding mRNAs.

To extend the analysis of the reliability of the procedures for isolating 5' ESTs from ESTs in the database, a similar analysis was conducted using a database composed of human mRNA sequences extracted from GenBank database release 97 for comparison. The 5' ends of more than 85% of 5' ESTs derived from mRNAs included in the GenBank database were located close to the 5' ends of the known sequence. As some of the mRNA sequences available in the GenBank database are deduced from genomic sequences, a 5' end matching with these sequences will be counted as an internal match. Thus, the method used here underestimates the yield of ESTs including the authentic 5' ends of their corresponding mRNAs.

**EXAMPLE 10**Calculation of Novelty Indices for 5'EST Libraries

In order to evaluate the novelty of 5'EST libraries, the following analysis was performed. For  
5 each sequenced 5'EST library, the sequences were clustered by the 5' end. Each sequence in the library was compared to the others and the longest sequence found in the cluster was used as representative of the group. A novelty rate (NR) was then defined as:  $NR = 100 \times (\text{Number of new unique sequences found in the library} / \text{Total number of sequences from the library})$ . Typically, novelty rating ranged between 10% and 41% depending on the tissue from which the 5'EST library was obtained. For most of  
10 the libraries, the random sequencing of 5' EST libraries was pursued until the novelty rate reached 20%.

**EXAMPLE 11**Generation of Consensus Contigated 5' ESTs

Since the cDNA libraries made above include multiple 5' ESTs derived from the same mRNA,  
15 overlapping 5'ESTs may be assembled into continuous sequences. The following method describes how to efficiently align multiple 5'ESTs in order to yield not only consensus contigated 5'EST sequences for mRNAs derived from different genes but also consensus contigated 5'EST sequences for different mRNAs, so called variants, transcribed from the same gene such as alternatively spliced mRNAs.

The whole set of sequences was first partitioned into small clusters containing sequences  
20 which exhibited perfect matches with each other on a given length and which derived from a small number of different genes. Some 5'EST sequences, so called singletons, were not aligned using this approach because they were not homologous to any other sequence.

Thereafter, all variants of a given gene were identified in each cluster using a proprietary software. 5'EST sequences belonging to the same variant were then contigated and consensus  
25 contigated 5'EST sequences generated for each variant. All consensus contigated 5' EST sequences were subsequently compared to the whole set of individual 5'EST sequences used to obtained them.

If desired, the consensus contigated 5'EST sequences may be verified by identifying clones in nucleic acid samples derived from biological tissues, such as cDNA libraries, which hybridize to the probes based on the sequences of the consensus contigated 5'ESTs using any methods described  
30 herein and sequencing those clones.

Application of this alignment method to a selected set of 5'ESTs free from endogenous contaminants and uninformative sequences, and following the masking of repeats, yielded consensus contigated 5'EST sequences or variants of clustered genes encompassing many individual 5'ESTs. Both non aligned 5'ESTs, *i.e.* singletons, and consensus contigated 5'ESTs were then compared to  
35 already known sequences and those sequences matching human mRNA sequences were eliminated from further analysis.

## EXAMPLE 12

Identification of Open Reading Frames in 5' ESTs

Subsequently, consensus contigated 5'ESTs and 5'ESTs were screened to identify those having an open reading frame (ORF).

5 Such open reading frames were simply defined as uninterrupted nucleic acid sequences longer than 45 nucleotides and beginning with an ATG codon.

Alternatively, the nucleic acid sequence was first divided into several subsequences which coding propensity was evaluated separately using one or several different methods known to those skilled in the art such as the evaluation of N-mer frequency and its variants (Fickett and Tung, 10 *Nucleic Acids Res*;20:6441-50 (1992)) or the Average Mutual Information method (Grosse *et al*, International Conference on Intelligent Systems for Molecular Biology, Montreal, Canada. June 28-July 1, 1998). Each of the scores obtained by the techniques described above were then normalized by their distribution extremities and then fused using a neural network into a unique score that represents the coding probability of a given subsequence. The coding probability scores obtained for 15 each subsequence, thus the probability score profiles obtained for each reading frame, was then linked to the initiation codons present on the sequence. For each open reading frame, defined as a nucleic acid sequence beginning with an ATG codon, an ORF score was determined. Preferably, this score is the sum of the probability scores computed for each subsequence corresponding to the considered ORF in the correct reading frame corrected by a function that negatively accounts for 20 locally high score values and positively accounts for sustained high score values. The most probable ORF with the highest score was selected.

In some embodiments, nucleic acid sequences encoding an "incomplete ORF", as referred therein, namely an open reading frame in which a start codon has been identified but no stop codon has been identified, were obtained.

25 In other embodiments, nucleic acid sequences encoding a "complete ORF", as used therein, namely an open reading frame in which a start codon and a stop codon have been identified, are obtained.

In a preferred embodiment, open reading frames encoding polypeptides of at least 50 amino acids were obtained.

30 To confirm that the chosen ORF actually encodes a polypeptide, the consensus contigated 5'EST or 5'EST may be used to obtain an extended cDNA using any of the techniques described therein, and especially those described in Examples 19 and 20. Then, such obtained extended cDNAs may be screened for the most probable open reading frame using any of the techniques described therein. The amino acid sequence of the ORF encoded by the consensus contigated 5'EST or 5'EST may then be 35 compared to the amino acid sequence of the ORF encoded by the extended cDNA using any of the algorithms and parameters described therein in order to determine whether the ORF encoded by the extended cDNA is basically the same as the one encoded by the consensus contigated 5'EST or 5'EST.



Alternatively, to confirm that the chosen ORF actually encodes a polypeptide, the consensus contiguated 5'EST or 5'EST may be used to obtain an extended cDNA using any of the techniques described therein, and especially those described in Examples 19 and 20. Such an extended cDNA may then be inserted into an appropriate expression vector and used to express the polypeptide encoded by the extended cDNA as described therein. The expressed polypeptide may be isolated, purified, or enriched as described therein. Several methods known to those skilled in the art may then be used to determine whether the expressed polypeptide is the one actually encoded by the chosen ORF, therein referred to as the expected polypeptide. Such methods are based on the determination of predictable features of the expressed polypeptide, including but not limited to its amino acid sequence, its size or its charge, and the comparison of these features to those predicted for the expected polypeptide. The following paragraphs present examples of such methods.

One of these methods consists in the determination of at least a portion of the amino acid sequence of the expressed polypeptide using any technique known to those skilled in the art. For example, the amino-terminal residues may be determined using techniques either based on Sanger's technique of acid hydrolysis of a polypeptide which N-terminal residue has been covalently labeled or using techniques based on Edman degradation of polypeptides which N-terminal residues are sequentially labeled and cleaved from the polypeptide of interest. The amino acid sequence of the expressed polypeptide may then be compared to the one predicted for the expected polypeptide using any algorithm and parameters described therein.

Alternatively, the size of the expressed polypeptides may be determined using techniques familiar to those skilled in the art such as Coomassie blue or silver staining and subsequently compared to the size predicted for the expected polypeptide. Generally, the band corresponding to the expressed polypeptide will have a mobility near that expected based on the number of amino acids in the open reading frame of the extended cDNA. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Alternatively, specific antibodies or antipeptides may be generated against the expected polypeptide as described in Example 34 and used to perform immunoblotting or immunoprecipitation studies against the expressed polypeptide. The presence of a band in samples from cells containing the expression vector with the extended cDNA which is absent in samples from cells containing the expression vector encoding an irrelevant polypeptide indicates that the expected polypeptide or portion thereof is being expressed. Generally, the band corresponding to the expressed polypeptide will have a mobility near that expected based on the number of amino acids in the open reading frame of the extended cDNA. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage

35

### EXAMPLE 13

#### Identification of Potential Signal Sequences in 5' ESTs

The 5'ESTs or consensus contigated 5'ESTs found to encode an ORF were then searched to identify potential signal motifs using slight modifications of the procedures disclosed in Von Heijne, *Nucleic Acids Res.* 14:4683-4690, 1986. Those sequences encoding a 15 amino acid long stretch with a score of at least 3.5 in the Von Heijne signal peptide identification matrix were considered to possess a signal sequence. Those nucleic acid sequences which match a known human mRNA or EST sequence and have a 5' end located downstream of the known 5' end, preferably by more than 20 nucleotides, were excluded from further analysis. The remaining nucleic acids having signal sequences therein were included in a database called SignalTag™.

10

#### EXAMPLE 14

##### Confirmation of Accuracy of Identification of Potential Signal Sequences in 5' ESTs

The accuracy of the above procedure for identifying signal sequences encoding signal peptides was evaluated by applying the method to the 43 amino acids located at the N terminus of all human SwissProt proteins. The computed Von Heijne score for each protein was compared with the known characterization of the protein as being a secreted protein or a non-secreted protein. In this manner, the number of non-secreted proteins having a score higher than 3.5 (false positives) and the number of secreted proteins having a score lower than 3.5 (false negatives) could be calculated.

Using the results of the above analysis, the probability that a peptide encoded by the 5' region of the mRNA is in fact a genuine signal peptide based on its Von Heijne's score was calculated based on either the assumption that 10% of human proteins are secreted or the assumption that 20% of human proteins are secreted. The results of this analysis are shown in Figure 2.

Using the above method of identification of secretory proteins, 5' ESTs of the following polypeptides known to be secreted were obtained: human glucagon, gamma interferon induced monokine precursor, secreted cyclophilin-like protein, human pleiotropin, and human biotinidase precursor. Thus, the above method successfully identified those 5' ESTs which encode a signal peptide.

To confirm that the signal peptide encoded by the 5' ESTs or consensus contigated 5' ESTs actually functions as a signal peptide, the signal sequences from the 5' ESTs or consensus contigated 5' ESTs may be cloned into a vector designed for the identification of signal peptides. Such vectors are designed to confer the ability to grow in selective medium only to host cells containing a vector with an operably linked signal sequence. For example, to confirm that a 5' EST or consensus contigated 5' EST encodes a genuine signal peptide, the signal sequence of the 5' EST or consensus contigated 5' EST may be inserted upstream and in frame with a non-secreted form of the yeast invertase gene in signal peptide selection vectors such as those described in U.S. Patent No. 5,536,637. Growth of host cells containing signal sequence selection vectors with the correctly inserted 5' EST or consensus contigated 5' EST signal sequence confirms that the 5' EST or consensus contigated 5' ESTs encodes a genuine signal peptide.

Alternatively, the presence of a signal peptide may be confirmed by cloning the extended cDNAs obtained using the ESTs or consensus contigated 5' ESTs into expression vectors such as pXT1 as described below, or by constructing promoter-signal sequence-reporter gene vectors which encode fusion proteins between the signal peptide and an assayable reporter protein. After introduction of these vectors into a suitable host cell, such as COS cells or NIH 3T3 cells, the growth medium may be harvested and analyzed for the presence of the secreted protein. The medium from these cells is compared to the medium from control cells containing vectors lacking the signal sequence or extended cDNA insert to identify vectors which encode a functional signal peptide or an authentic secreted protein.

### EXAMPLE 15

#### Analysis of the Sequences of the Invention

The set of the nucleic acid sequences of the invention (SEQ ID NOs. 24-811 and 1600-1622) was obtained as described in Example 11. Subsequently, the most probable open reading frame was determined and signal sequences were searched, as described in Examples 12 and 13, for all sequences of the invention.

The nucleotide sequences of the SEQ ID NOs. 24-811 and 1600-1622 and the polypeptide sequences encoded by SEQ ID NOs. 24-811 (*i.e.* polypeptide sequences of SEQ ID NOs. 812-1599) are provided in the appended sequence listing which structure is as follows.

SEQ ID NOs. 24-728 are nucleic acids having an incomplete ORF which encodes a signal peptide. The locations of the incomplete ORFs and sequences encoding signal peptides are listed in the accompanying Sequence Listing. In addition, the von Heijne score of the signal peptide computed as described in Example 13 is listed as the "score" in the accompanying Sequence Listing. The sequence of the signal-peptide is listed as "seq" in the accompanying Sequence Listing. The "/" in the signal peptide sequence indicates the location where proteolytic cleavage of the signal peptide occurs to generate a mature protein.

SEQ ID NOs. 729-765 are nucleic acids having an incomplete ORF in which no sequence encoding a signal peptide has been identified to date. However, it remains possible that subsequent analysis will identify a sequence encoding a signal peptide in these nucleic acids. The locations of the incomplete ORFs are listed in the accompanying Sequence Listing.

SEQ ID NOs. 766-792 are nucleic acids having a complete ORF which encodes a signal peptide. The locations of the complete ORFs and of the signal peptides, the von Heijne score of the signal peptide, the sequence of the signal-peptide and the proteolytic cleavage site are indicated as described above.

SEQ ID NOs. 793-811 are nucleic acids having a complete ORF in which no sequence encoding a signal peptide has been identified to date. However, it remains possible that subsequent analysis will

identify a sequence encoding a signal peptide in these nucleic acids. The locations of the complete ORFs are listed in the accompanying Sequence Listing.

SEQ ID NOs. 812-1516 are "incomplete polypeptide sequences" which include a signal peptide. "Incomplete polypeptide sequences" are polypeptide sequences encoded by nucleic acids in which a start  
5 codon has been identified but no stop codon has been identified. These polypeptides are encoded by the nucleic acids of SEQ ID NOs. 24-728. The location of the signal peptide, the von Heijne score of the signal peptide, the sequence of the signal-peptide and the proteolytic cleavage site are indicated as described above.

SEQ ID NOs. 1517-1553 are incomplete polypeptide sequences in which no signal peptide has  
10 been identified to date. However, it remains possible that subsequent analysis will identify a signal peptide in these polypeptides. These polypeptides are encoded by the nucleic acids of SEQ ID NOs. 729-765.

SEQ ID NOs. 1554-1580 are "complete polypeptide sequences" which include a signal peptide. "Complete polypeptide sequences" are polypeptide sequences encoded by nucleic acids in which a start  
15 codon and a stop codon have been identified. These polypeptides are encoded by the nucleic acids of SEQ ID NOs. 766-792. The location of the signal peptide, the von Heijne score of the signal peptide, the sequence of the signal-peptide and the proteolytic cleavage site are indicated as described above..

SEQ ID NOs. 1581-1599 are complete polypeptide sequences in which no signal peptide has been identified to date. However, it remains possible that subsequent analysis will identify a signal  
20 peptide in these polypeptides. These polypeptides are encoded by the nucleic acids of SEQ ID NOs. 793-811.

SEQ ID NOs. 1600-1622 are nucleic acid sequences in which no open reading frame has been conclusively identified to date. However, it remains possible subsequent analysis will identify an open reading frame in these nucleic acids.

25 In the accompanying Sequence Listing, all instances of the symbol "n" in the nucleic acid sequences mean that the nucleotide can be adenine, guanine, cytosine or thymine. In some instances the polypeptide sequences in the Sequence Listing contain the symbol "Xaa." These "Xaa" symbols indicate either (1) a residue which cannot be identified because of nucleotide sequence ambiguity or (2) a stop codon in the determined sequence where applicants believe one should not exist (if the sequence  
30 were determined more accurately). In some instances, several possible identities of the unknown amino acids may be suggested by the genetic code.

In the case of secreted proteins, it should be noted that, in accordance with the regulations governing Sequence Listings, in the appended Sequence Listing, the full protein (*i.e.* the protein containing the signal peptide and the mature protein) extends from an amino acid residue having a  
35 negative number through a positively numbered C-terminal amino acid residue. Thus, the first amino acid of the mature protein resulting from cleavage of the signal peptide is designated as amino acid

number 1, and the first amino acid of the signal peptide is designated with the appropriate negative number.

If one of the nucleic acid sequences of SEQ ID NOs. 24-811 and 1600-1622 are suspected of containing one or more incorrect or ambiguous nucleotides, the ambiguities can readily be resolved by resequencing a fragment containing the nucleotides to be evaluated. If one or more incorrect or ambiguous nucleotides are detected, the corrected sequences should be included in the clusters from which the sequences were isolated, and used to compute other consensus contigated sequences on which other ORFs would be identified. Nucleic acid fragments for resolving sequencing errors or ambiguities may be obtained from deposited clones or can be isolated using the techniques described herein.

Resolution of any such ambiguities or errors may be facilitated by using primers which hybridize to sequences located close to the ambiguous or erroneous sequences. For example, the primers may hybridize to sequences within 50-75 bases of the ambiguity or error. Upon resolution of an error or ambiguity, the corresponding corrections can be made in the protein sequences encoded by the DNA containing the error or ambiguity. The amino acid sequence of the protein encoded by a particular clone can also be determined by expression of the clone in a suitable host cell, collecting the protein, and determining its sequence.

In addition, if one of the sequences of SEQ ID NOs. 812-1599 is suspected of containing a truncated ORF as the result of a frameshift in the sequence, such frameshifting errors may be corrected by combining the following two approaches. The first one involves thorough examination of all double predictions, *i.e.* all cases where the probability scores for two ORFs located on different reading frames are high and close, preferably different by less than 0.4. The fine examination of the region where the two possible ORFs overlap may help to detect the frameshift. In the second approach, homologies with known proteins are used to correct suspected frameshifts.

Of the identified clusters, some were shown to be multivariant, *i.e.* to contain several variants of the same gene. Table I gives for each of the multivariant clusters named by its internal reference (first column), the list of all variant consensus contigated 5'ESTs (second column), each being represented by a different sequence identification number.

TABLE I

Cluster Internal Reference	SEQ ID NOs of Variants
C1	687, 791
C2	744, 798
C3	640, 811
C4	59, 66
C5	84, 97

C6	287, 289
C7	286, 775, 777
C8	762, 768
C9	783, 784
C10	80, 1603
C11	655, 736
C12	805, 806

Table II provides a list preferred polynucleotide fragments which are derivatives of the consensus contigated 5'ESTs. As used herein the term "polynucleotide described in Table II" refers to the all of the preferred polynucleotide fragments defined in Table II in the following manner. The fragments are referred to by their SEQ ID numbers in the first column. The preferred polynucleotide fragments are then defined by a range of nucleotide positions from the SEQ IDs of the consensus contigated 5'ESTs as indicated in the second column entitled "positions of preferred fragments." The preferred polynucleotide fragments correspond to the individual 5'ESTs aligned to obtain the consensus contigated 5'EST and to those filed in the priority documents. The third column entitled "variant nucleotides" describes the nucleotide sequence variations observed between the consensus contigated 5'EST and preferred nucleic acid fragments as follows:

A) Substitutions in the sequence of a consensus contigated 5'EST to derive a preferred polynucleotide fragment are denoted by an "S", followed by a number indicating the first nucleotide position in a specific SEQ ID to be substituted in a string of substituted nucleotides or the position of the substituted nucleotide in the case of a single substituted nucleotide. Then there is a coma followed by one or more lower case letters indicating the identity of the nucleotide(s) occurring in the substituted position(s). For example, SEQ ID NO: 3401; Position of preferred fragments: 1-250; Variant nucleotides S45,atc would indicate that a preferred polynucleotide fragment had the sequence of positions 1 to 250 of SEQ ID NO. 3401, except that the nucleotides at positions 45, 46, and 47 were substituted with A, T, and C, respectively, in the preferred polynucleotide as compared with the sequence of SEQ ID No. 3401.

B) Insertions in the sequence of a consensus contigated 5'EST to derive a preferred polynucleotide fragment are denoted by an "I", followed by a number indicating the nucleotide position in a specific SEQ ID after which a string of nucleotides is inserted or the position after which the nucleotide is inserted in the case of a single inserted nucleotide. Then there is a coma followed by one or more lower case letters indicating the identity of the nucleotide(s) occurring in the inserted position(s). For example, SEQ ID NO: 7934; Position of preferred fragments: 1-500; Variant nucleotides: I36,gataca would indicate that a preferred polynucleotide fragment had the sequence of positions 1 to 500 of SEQ ID NO. 7934, except that after the nucleotides at position 36 a GATACA string of nucleotides is inserted in the preferred polynucleotide as compared with the sequence of SEQ ID No. 7934.

C) Deletions in the sequence of a consensus contigated 5'EST to derive a preferred nucleic acid fragment are denoted by an "D", followed by a number indicating the first nucleotide position in a specific SEQ ID to be deleted in a string of deleted nucleotides or the position of the deleted nucleotide in the case of a single deleted nucleotide. Then there is a comma followed by number indicating the number of nucleotide(s) deleted from the sequence provided in the sequence ID. For example, SEQ ID NO: 5398; Position of preferred fragments: 56-780; Variant nucleotides D114,5 would indicate that a preferred polynucleotide fragment had the sequence of positions 56 to 780 of SEQ ID NO. 5398, except that the nucleotides in positions 114 to 118 had been deleted in the preferred polynucleotide as compared with the sequence of SEQ ID No. 5398.

The present invention encompasses isolated, purified, or recombinant nucleic acids which consist of, consist essentially of, or comprise a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, or 500 nucleotides in length, to the extent that a contiguous span of these lengths is consistent with the lengths of the particular polynucleotide, of a polynucleotide described in Table II, or a sequence complementary thereto, wherein said polynucleotide described in Table II is selected individually or in any combination from the polynucleotides described in Table II. The present invention also encompasses isolated, purified, or recombinant nucleic acids which consist of or consist essentially of a polynucleotide described in Table II, or a sequence complementary thereto, wherein said polynucleotide is selected individually or in any combination from the polynucleotides described in Table II. The present invention further encompasses isolated or purified polypeptides which consist of, consist essentially of, or comprise a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, or 100 amino acids encoded by a polynucleotide described in Table II.

Table II

SEQ ID NO.	Positions of Preferred Fragments	Variant nucleotides
35	1-423	S124, s; I135, a; S293, w; I363, a; S377, r; D424, 15
41	1-427	I117, m; S120, r; S124, g; D373, l; S376, b; S378, b; I427, gggg; D428, 109
43	1-276	S114, m; S118, rg; S123, r; S139, nr; I142, t; D148, l; D152, l; I228, t; I276, gg; D277, 136
45	126-420	D1, 125; I420, ggg; D421, 100
46	1-255	S139, r; I145, r; S146, mm; S150, ar; S254, g; D256, 128
48	4-437	D1, 3; S49, a; S55, g; S79, a; S90, a; I437, tctctg
59	1-471	S26, a; S44, t; S48, t; S109, a; S191, t; S200, gc; S203, a; S210, g; S237, a; S240, g; S255, a; S272, a; S277, a; S279, a; S284, t; S297, g; S305, g; S316, a; I471, ggtca
66	1-428	I428, tactgggg

82	1-399	S251, t; S277, d; I399, aagccggg
84	5-488	D1, 4; S210, g; S293, a; S325, g; S339, a; S348, g; S353, g; S395, g; I488, cacca
93	1-508	I508, gattt
96	26-315	D1, 25; S28, a; S62, c; I315, cagatgg
97	4-460	D1, 3; S19, g; S31, g; S114, gt; S118, a; S123, tc; S127, c; S132, a; S186, g; S190, c; S203, t; S210, g; S232, c; I460, acgtt
105	1-281	S273, a; I281, g; D282, 211
114	10-315	I0, t; D1, 9; S91, m; S267, n; S276, w; S292, h; S295, m; I315, tggg; D316, 19
118	1-145	S57, d; S126, d; I145, ccctc
120	2-348	D1, 1; S104, t; I348, g; D349, 38
121	1-190	I121, c; I190, ccctt
123	1-353	I117, m; I186, w; S187, y; I353, caccgggg
124	1-249	I249, ggrvgggg
125	114-375	D1, 113; S206, wn; I231, a; I375, ccctagg
126	1-437	S297, cc; S307, tg; S312, a; S318, g; S341, a; S351, t; S353, g; S383, c; S387, a; D404, 1
136	82-428	D1, 81; I428, aaagtg
139	1-268	I268, gggaaggg
148	6-405	D1, 5; I405, ggtgt
159	1-230	S227, ta; I230, ccctggg
165	3-256	I0, tat; D1, 2; I17, c; S18, t; S111, d; I115, t; S123, r; I256, aaggcggg
170	1-280	I103, t; S104, c; I111, t; I280, cgctcggg
194	1-215	S50, s; S186, sn; S199, k; I215, gcagcggg
213	1-158	S128, m; I132, w; S143, d; I158, tgcccggg
223	3-431	D1, 2; S28, s; S79, c; S82, s; S308, nr; S328, nb; I431, ccggc
247	1-359	I76, gttt; I359, tccctgg
258	1-236	S72, r; S81, g; S197, s; I205, ss; S232, k; I236, acttcggg
264	5-283	D1, 4; S64, g; S122, m; S134, yy; I137, c; I151, t; I283, gttgc
269	1-143	S111, s; I143, ggggcggg
286	5-207	D1, 4; S204, a; S206, c; I207, gg; D208, 567
287	1-277	S114, r; I125, t; S131, ag; S256, tg; S259, tt; S262, at; S267, t; S269, c; S273, c; I277, ccggg; D278, 337
289	69-416	D1, 68; I416, agccaggg
289	1-278	S114, r; I125, t; S131, ag; S277, c; I278, cggg; D279, 138
292	20-254	D1, 19; I254, aaagagg
293	1-414	I414, tagcag
300	1-285	S16, m; S67, y; I285, baccacggg; D286, 1
349	23-431	D1, 22; I118, a; S214, y; I431, caactgg
350	3-386	D1, 2; S42, w; I263, c; I386, gggat
368	3-446	D1, 2; I446, tctct
385	1-193	I35, t; I108, t; I134, r; S135, a; S137, r; S143, w; I178, c; I193, gagcgggg
411	6-391	D1, 5; S17, r; S27, t; S334, y; D392, 244
412	1-185	S49, s; S127, s; I185, gctggg; D186, 150



415	2-229	D1, 1; S3, a; I229, caaatggg
435	1-386	S4, s; I386, ccggg
436	4-472	D1, 3; S61, sa; D238, 1; S239, s; I472, agtgtgg
437	1-340	I340, ggg; D341, 129
441	1-409	S109, smag; I409, cgcacggg
454	1-492	S72, nn; S115, t; S121, bwy; S181, yn; I492, gagtc
455	1-177	I14, w; I16, a; I177, gagctggg
459	1-311	S39, n; S74, rg; I311, accatggg
460	1-425	I425, agtac
461	5-420	D1, 4; I420, tcgtc
481	1-429	I10, w; S262, d; S333, n; I429, ctccaggg
489	1-414	D72, 1; S117, n; S396, d; I414, ggaca
496	1-215	I215, ttctcggg
501	1-430	S275, n; I430, aggat
502	91-413	D1, 90; I413, aaacgggg
504	21-420	D1, 20; S47, w; S83, n; I280, n; S281, na; S292, v; S314, sm; S368, ww; S373, w; I420, cccca
505	18-457	D1, 17; D36, 1; S182, g; S273, n; S283, a; S416, bh; I457, ctcca
514	1-303	I303, accca
515	1-455	S11, t; I12, n; S30, r; S256, wr; I333, t; I455, cataa
517	24-453	D1, 23; I453, agagcggg
519	1-275	I119, gt; S125, w; I129, w; S133, k; S137, k; S167, k; I275, gcccc
522	1-313	I313, agcgtggg
526	4-366	I0, t; D1, 3; I366, ggccccggg
530	1-434	S328, g; I434, aagat
535	1-379	S128, g; S162, m; D380, 5
561	2-341	D1, 1; I341, raagagg
568	1-246	I118, g; S137, g; I246, aaaccggg
570	1-207	I207, tttt
576	1-288	I34, c; I288, cccgtgg
588	1-390	S218, a; S224, k; S314, dh; S358, s; D376, 1; I390, atg; D391, 23
597	31-274	D1, 30; S49, n; I274, tccatgg
606	1-354	I141, g; D174, 1; S229, rr; D355, 72
627	1-415	S7, a; I415, cattt
634	1-178	D179, 212
640	6-428	D1, 5; D429, 79
641	64-483	D1, 63; I165, d; D183, 1; S185, y; S253, t; D279, 2; S416, a; I483, atata
655	1-280	S58, c; I84, g; S88, k; S204, ac; S244, g; S247, g; I280, ggg; D281, 90
672	34-489	D1, 33; S316, k; S331, k; S333, w; S486, g; S488, c; D490, 4
687	116-473	D1, 115; S142, n; I473, cctcgggg
697	1-202	S142, s; S144, sr; S148, d; S152, d; I155, a; I164, a; S174, k; I202, gcc; D203, 291
708	8-384	D1, 7; S104, b; I384, gaaaa
710	1-167	S40, k; S49, db; I167, tatct

722	1-191	I125, c; I191, tttt
723	1-316	I316, aggg; D317, 157
729	15-373	D1, 14; S139, t; I373, cgcag; D374, 99
730	29-372	D1, 28; I155, g; S192, ka; S333, d; I372, m; D373, 93
731	1-290	S10, kk; S30, b; S32, t; S92, t; S197, dy; S278, g; I290, aggg; D291, 55
732	8-277	D1, 7; I113, a; S127, w; I131, s; S132, r; S156, w; S160, r; S211, n; S215, w; I247, a; D278, 121
733	20-375	D1, 19; S306, sbs; I325, h; S326, nr; S338, ywd; S344, v; I375, aggg; D376, 68
734	1-359	D66, 1; D360, 14
735	25-322	D1, 24; S30, r; I193, a; I322, ccaaggg
736	9-181	D1, 8; S58, g; I181, aactaggg
737	1-160	S97, ta; I160, aggtc
738	1-227	D228, 7
739	45-514	D1, 44; S178, s; I182, c; S436, dmn; S461, v; S476, c; S506, t; D515, 75
740	11-388	D1, 10; I388, cgacaggg
741	1-478	S118, s; S125, a; I126, s; S134, k; S421, vn; I478, aatsc
742	217-553	I0, tt; D1, 216; S286, r; S294, m; S311, r; S317, s; S338, r; S442, dm; S469, h; S476, r; S485, s; S491, w; I495, ht; S496, v; S513, r; D521, 1; S536, m; D554, 199
743	1-459	I11, s; S258, m; I270, m; I304, c; I308, amta; S313, c; S438, v; I459, agggag
744	25-316	D1, 24; S315, g; D317, 95
745	21-283	D1, 20; I40, g; S41, c; D123, 1; S181, sr; S227, r; I283, ccgcg; D284, 121
746	1-256	D257, 173
747	1-179	S134, w; S138, w; S140, kt; I179, cacca
748	1-235	S46, t; I72, t; S189, cc; S222, c; D236, 148
749	2-370	D1, 1; S32, cg; D144, 1; S341, g; D371, 76
750	18-410	I0, aag; D1, 17; I410, aatcc
751	22-355	D1, 21; D148, 1; S150, c; S152, a; S313, n; D356, 181
752	1-139	S50, t; I118, g; I139, ccct
753	1-189	S26, r; S115, s; I121, r; S122, r; S128, s; S143, r; I146, w; S156, r; D190, 4
754	1-395	S212, wd; I395, cggca
755	19-460	D1, 18; S26, c; S156, a; S253, n; I460, tagaagg
756	2-142	D1, 1; I106, gc; S107, t; S110, c; I142, ccaccggg
757	28-296	D1, 27; I119, s; I122, t; S128, s; S255, t; S267, m; D297, 66
758	11-368	D1, 10; I200, g; S201, c; S281, d; S317, c; I368, ccatcggg
759	19-452	D1, 18; S421, w; I452, a
760	25-175	D1, 24; S34, yk; I175, ccggg; D176, 120
761	1-212	I212, cactcggg
762	1-374	S320, s; S349, a; D375, 249
763	8-152	D1, 7; I152, acggg; D153, 109

764	1-160	I127, g; I145, g; I160, cgcccggg
765	137-313	D1, 136; S272, m; I279, s; S310, t; I313, ggg; D314, 203
766	1-320	S278, ag; S281, cagacc; S288, ta; S291, caag; S296, c; S317, m; I320, cggg; D321, 306
767	6-336	I0, aa; D1, 5; S149, w; S245, y; D337, 137
768	1-374	S320, s; D375, 299
769	53-435	D1, 52; S59, b; S344, nnkw; D436, 104
770	24-448	D1, 23; S25, g; S411, w; S416, m; D449, 31
771	1-370	S3, c; S180, m; S275, r; D371, 122
772	1-388	I299, c; S326, c; D389, 8
773	1-143	S18, c; S66, a; I143, ggg; D144, 274
774	1-347	S194, a; S205, c; I347, ggg; D348, 107
775	5-207	D1, 4; S111, tg; S158, g; S171, c; S191, a; S204, a; S206, c; I207, gg; D208, 324
776	1-368	I200, c; S201, a; S291, ta; I332, c
777	5-207	D1, 4; S204, a; S206, c; I207, gg; D208, 262
778	39-342	D1, 38; S184, r; D343, 126
779	4-360	D1, 3; S13, m; S15, c; S22, s; S24, m; S48, r; S56, s; S335, c; S345, rs; I360, ggg; D361, 119
780	1-472	I347, c; D473, 32
781	116-426	D1, 115; S219, m; S424, g; D427, 118
782	1-391	S386, k; D392, 64
783	1-453	D109, l; S110, y; S125, y; I128, g; S132, k; I453, ctctc
784	29-494	D1, 28; S72, r; D495, 93
785	99-461	D1, 98; S218, r; I461, gaccgggg
786	2-465	D1, 1; S8, y; S388, s; I398, g; S400, t; S403, at; S417, g; D466, 24
787	28-271	D1, 27; S99, t; S230, c; S266, ga; S269, c; I271, g; D272, 126
788	1-285	D280, l; I285, g; D286, 310
789	1-209	S205, c; D210, 150
790	51-297	D1, 50; I297, ggggg; D298, 539
791	113-327	D1, 112; S218, g; I226, g; D280, l; I327, cgagg; D328, 224
792	17-218	D1, 16; S58, t; S217, t; I218, gggg; D219, 219
793	11-92	D1, 10; S91, c; I92, a; D93, 258
794	9-431	D1, 8; I431, taagt
795	30-341	D1, 29; I341, a; D342, 175
796	1-442	S17, w; S19, wr; D35, l; S134, t; S264, n; S322, nr; S369, s; S420, s; S422, y; I442, tcctcggg
797	1-420	S136, c; S150, c; I245, ccc; I420, ggagtg
798	25-316	D1, 24; S315, g; D317, 97
799	1-344	D345, 57
800	7-465	D1, 6; S59, k; S146, a; S186, krn; I465, gttca
801	121-422	D1, 120; I269, c; S419, cc; I422, gg; D423, 207
802	46-477	D1, 45; S132, bn; I477, actac
803	15-467	D1, 14; S45, k; S65, t; S418, ys; D452, l; D468, 119
804	1-341	S42, t; S97, d; S326, gtg; S331, tgt; S336, a;

		S338, c; I341, cccccggg; D342, 218
805	2-409	D1, 1; S334, d; I409, aggg; D410, 161
806	5-384	D1, 4; I384, actaa
807	1-301	S113, a; S117, c; S123, t; D128, 1; D134, 1; S282, g; S284, a; I301, gacggagggg; D302, 70
808	2-314	D1, 1; S306, g; I314, ggg; D315, 121
809	1-394	S53, g; S228, n; S272, vk; I301, g; I358, m; S368, nb; S375, w; I383, mm; I388, yt; I394, nhaccggg
810	6-205	I0, a; D1, 5; I141, t; I205, ggg; D206, 630
811	6-270	D1, 5; I270, gggg; D271, 115
1600	1-247	S45, m; S114, k; I122, m; S123, yc; S158, rr; S221, k; I247, ccccaggg
1601	1-225	S109, bm; S195, m; I225, tgcacggg
1602	23-245	D1, 22; D138, 1; S139, s; S242, t; S244, g; I245, g; D246, 13
1603	1-303	S71, c; D277, 1; I303, ggagggg; D304, 38
1604	1-242	S47, w; S50, c; S81, h; S85, d; S91, k; S106, r; I242, tgtggg; D243, 50
1605	2-225	D1, 1; S20, k; S91, c; I225, ggg; D226, 132
1606	15-293	D1, 14; S156, g; S193, g; I200, t; I293, acaaaggg
1607	1-361	S323, c; I361, ccca
1608	1-151	I151, taagggg; D152, 154
1609	1-242	S55, s; I135, a; S152, h; I242, cagtaggg
1610	1-196	I151, w; S190, k; I196, cctgtgg
1611	1-228	S115, k; S174, rk; I228, cgtttggg
1612	1-221	S108, v; I221, tgatcggg
1613	1-281	I66, w; I137, a; D282, 79
1614	1-171	S53, k; S76, k; I80, k; S81, kw; S86, r; S92, k; S126, k; I171, gccgagg
1615	2-193	D1, 1; S67, c; I121, s; S122, mm; S126, g; S130, r; S146, r; S156, gm; I193, cctca
1616	1-349	S251, ww; S259, rs; S275, k; I279, w; S285, y; S292, y; I320, m; I331, m; I338, w; I341, s; I349, accccggg
1617	1-129	I118, t; D130, 26
1618	1-184	D9, 1; D185, 1
1619	1-169	I122, t; I169, gcccgagg
1620	1-187	S106, k; S118, m; S122, cg; S132, k; D188, 59
1621	1-153	D125, 1; I131, tt; S152, t; I153, gg; D154, 127
1622	1-400	S43, s; I126, g; I129, y; S353, d; I400, tatat

### EXAMPLE 16

#### Categorization of 5' ESTs and Consensus Contigated 5'ESTs

The nucleic acid sequences of the present invention (SEQ ID NOs. 24-811 and 1600-1622) were  
5 grouped based on their homology to known sequences as follows. All sequences were compared to  
EMBL release 57 and daily releases available at the time of filing using BLASTN. All matches with a  
minimum of 25 nucleotides with 90% homology were retrieved and used to compute Tables IV and V.

In some embodiments, 5'ESTs or consensus contigated 5'ESTs nucleic acid sequence do not match any known vertebrate sequence nor any publicly available EST sequence, thus being completely new.

In other embodiments, 5'ESTs or consensus contigated 5'ESTs match a known sequence.

- 5 Tables III and IV gives for each sequence of the invention in this category referred to by its sequence identification number in the first column, the positions of their preferred fragments in the second column entitled "Positions of preferred fragments." As used herein the term "polynucleotide described in Table III" refers to the all of the preferred polynucleotide fragments defined in Table III in this manner, and the term "polynucleotide described in Table IV" refers to the all of the preferred polynucleotides fragments
- 10 defined in Table IV in this manner. The present invention encompasses isolated, purified, or recombinant nucleic acids which consist of, consist essentially of, or comprise a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, or 500 nucleotides in length, to the extent that a contiguous span of these lengths is consistent with the lengths of the particular polynucleotide, of a polynucleotide described in Table III or Table IV, or a sequence complementary thereto, wherein said
- 15 polynucleotide described in Table III or Table IV is selected individually or in any combination from the polynucleotides described in Table III or Table IV. The present invention also encompasses isolated, purified, or recombinant nucleic acids which consist of or consist essentially of a polynucleotide described in Table III or Table IV, or a sequence complementary thereto, wherein said polynucleotide is selected individually or in any combination from the polynucleotides described in Table III or Table IV.

20

Table III

SEQ ID NO	Positions of preferred fragments
24	1-251
25	1-83
28	227-276
29	1-27
30	130-242, 283-315, 365-461
32	314-399
33	89-321
34	1-38
35	1-52, 171-222
36	1-30, 408-441
37	1-138
39	115-140
40	1-97
41	1-112
42	1-177
46	1-38
48	376-400
51	400-466
54	1-259
55	189-320

56	265-457
58	246-469
59	81-123, 418-444
60	1-348
61	78-123, 418-457
62	386-439
63	1-214
64	109-297
65	1-370
66	92-428
68	1-180
69	165-259
70	1-178
71	1-27
72	1-179
73	1-65, 107-192
75	1-314
77	263-388
78	1-64
79	1-149
80	101-142, 302-380
82	1-192
83	1-398
85	1-290
86	1-118, 149-336
87	1-262
88	1-149
89	1-315
90	1-74
91	1-335, 364-423
92	1-316
93	338-508
94	179-321
95	219-402
96	26-315
97	348-460
98	1-230
99	391-467
101	214-336
102	1-289
103	1-383
104	1-211
105	1-36
106	1-126
107	1-49
108	294-336
109	1-128
111	1-154
112	407-441
113	1-80, 139-184
114	10-79
116	1-292
117	1-304

119	1-288
120	2-348
121	1-122
123	188-353
124	1-249
125	295-375
128	1-244
129	1-232
130	196-312
131	178-276
132	37-174
133	1-344
134	1-244
135	1-217
136	82-428
137	1-29, 103-155, 274-434
138	1-395
139	1-268
140	1-170
141	1-396
142	1-73, 227-357
143	1-159
144	1-433
145	61-116
146	1-71, 179-205
147	177-300
149	1-146
151	1-166
152	1-382
153	1-208
154	121-251
155	1-147
157	1-115
158	1-175
159	1-44, 80-230
160	1-346
161	1-277
162	1-235
163	1-34
164	1-195
165	19-78, 175-217
166	1-209
167	1-65
168	128-218
169	49-245
170	179-280
171	1-103
172	1-218
173	1-380
174	1-139
175	1-122
176	1-300
177	1-466

179	1-86
180	1-245
181	1-241
182	1-263
183	1-170
184	58-106, 399-443
185	1-427
186	1-365
187	1-260
188	1-172
189	1-150
190	161-271, 301-339
191	1-91
192	1-264
193	1-246
194	1-150
195	1-209
196	1-363
197	1-155
198	1-135
200	1-125
201	1-210
202	1-338
203	1-188
204	228-347
205	1-440
206	56-221
208	1-422
209	169-195
210	1-363
211	1-368
212	1-448
213	1-134
214	1-193
215	1-214
216	1-134
218	1-189
219	1-248
220	1-115
221	1-113
222	1-370
224	1-251
225	1-198
226	45-141
227	1-206
228	1-480
229	1-144
230	1-42, 281-351, 432-457
231	1-112
233	1-301
234	1-109
235	1-393
236	1-222



237	1-154
238	1-439
239	112-137
240	1-194
241	1-44
242	1-242
244	1-324
245	1-38, 217-280
246	1-60
247	77-359
248	1-236
249	1-342
250	80-382
251	1-303
252	62-259
253	1-165
254	1-328
255	1-320
256	1-305
257	1-181
258	116-174
259	1-265
260	1-272
261	1-62
263	1-371
266	1-274
267	1-342
268	364-427
269	31-143
270	1-79
271	1-121
272	229-292
273	1-158
274	1-113
275	1-254
276	1-333
277	1-130
278	1-184
279	1-265
280	1-188
281	1-177
282	1-336
283	1-294
284	1-171
285	1-297
288	1-42
290	1-170
292	20-155
294	1-334
295	1-375
296	1-226
297	1-232
299	40-139

300	1-285
301	1-242
302	1-136
303	1-175
304	1-493
305	1-214
306	89-458
307	1-328
308	1-380
309	1-236
310	1-357
311	1-470
312	1-187
313	1-159
315	1-162
316	1-404
317	1-450
318	1-395
319	1-257
320	56-325
321	1-201
322	1-159
323	1-420
324	1-210
325	1-192
326	88-181
327	1-185
328	128-210
330	1-223
331	1-362
332	1-89
334	1-188
335	1-115
336	1-300
337	1-307
338	1-123
339	1-297
340	1-34
341	1-44
342	1-37
343	141-169
344	1-112
345	1-235, 266-349
346	1-191
347	1-229
348	1-210
350	139-266
351	1-307
352	1-170
353	1-293
354	30-161, 192-331
355	1-93
356	1-178

357	1-107
358	1-29, 168-209
359	1-298
360	1-193
362	1-360
363	1-45, 100-212
364	39-170, 202-242
365	1-248
366	1-351
367	1-208
368	228-446
369	1-62
370	1-132
371	1-127
372	1-196
373	1-148
374	1-126
375	1-112
376	1-146
378	1-143
379	1-261
380	202-228
382	1-151
383	1-45
384	1-190, 250-456
385	1-55, 141-181
386	1-281
387	1-111
388	1-374
389	1-192
390	1-371
392	1-303
394	1-126
395	1-329
396	1-99
397	1-316
398	1-251
399	1-120
401	1-206
402	1-330
403	1-311
405	1-153
406	1-206
407	1-479
408	1-289
410	229-321
413	1-158
415	95-229
416	1-265
417	1-228
418	1-225
419	207-293
420	1-194

421	1-90
422	1-161
423	1-420
424	1-432
425	1-276, 309-419
426	1-232
427	1-81
428	1-96
429	1-165
431	1-58, 186-237, 327-354
433	1-65
434	1-83
435	1-386
436	405-447
438	1-106
439	45-105, 168-255, 284-447
441	1-409
442	1-320
443	1-256
444	1-284
445	1-240
446	1-149
447	1-360
448	1-123
449	1-94
450	1-302
452	1-349
453	1-270
454	1-492
455	17-105
456	1-102
457	1-108
458	1-285
459	1-311
460	1-191
461	312-420
462	1-257
463	1-117
464	1-142
466	1-235
467	1-29
468	1-41
469	1-438
470	1-131
471	1-211
472	1-150
473	1-352
474	1-141
476	1-232
478	1-201
479	1-151
480	1-104
481	7-429

482	1-385
486	1-226
488	1-296
489	1-72, 323-377
491	1-348
492	33-126
493	1-300
494	1-295
495	1-244
496	1-215
497	1-255
499	1-174, 384-474
500	1-50, 102-241
501	153-430
502	91-132
503	1-64
504	21-63, 356-420
505	37-68, 187-234
506	1-315
507	101-208
510	1-402
511	1-343
512	1-140, 170-246, 276-420
513	1-324
514	1-303
515	13-340
516	1-263, 293-360
518	1-245
519	111-275
520	62-182
521	1-218
523	1-502
524	1-118
525	1-276
526	223-366
527	1-428
528	297-342
529	1-244
530	1-88, 375-434
531	1-406
533	1-149
534	1-145
535	1-116
536	1-207
537	1-394
538	1-415
539	1-160
540	1-327
541	1-38, 73-396
542	1-247
543	1-221
544	1-375
545	1-376

546	1-109
547	1-160, 223-306
548	1-148
551	1-231
552	1-229
553	1-232
554	1-141
555	1-376
556	1-279
557	1-340
558	1-51
559	1-354
562	1-188
563	1-229
564	184-352
566	308-341
567	1-218
568	1-79
569	1-142
570	1-207
571	1-373
572	1-195
573	1-352
574	1-121
575	1-222
576	151-288
577	1-264
578	1-205
580	1-171, 273-328
581	1-356
582	1-239
583	1-144
584	1-282
585	1-338
586	1-436
588	1-380
589	1-60
590	1-178
592	1-66
593	1-215
594	1-161
596	1-407
597	31-83
598	1-417
599	1-329
600	1-311
601	1-61, 99-214
602	1-154, 197-463
603	135-269
604	1-351
605	1-195
608	1-357
609	1-201

612	1-176
613	1-342
615	1-272
616	1-114
617	1-46
618	1-208
619	1-257
620	1-28
621	1-26
622	1-221
623	1-432
624	1-233
625	1-26
627	1-43
628	1-318
629	1-170
630	1-196
631	248-339
632	1-433
633	1-154
634	1-41
635	1-137
636	1-172
637	1-253
638	1-185
639	1-206
641	334-483
642	1-309
643	1-75, 162-213
644	107-211
645	1-98
646	1-347
647	1-49, 81-143
648	1-232
649	74-133
650	1-37
651	1-276
652	1-170
653	1-178
654	1-121
656	1-197
657	1-246
659	1-197
660	116-172
661	1-411
662	1-146
663	1-65
664	1-182
665	1-320
666	1-273
667	1-149
668	1-122
670	1-160

671	1-137
673	1-263
674	1-263
675	1-107
677	1-441
678	134-191
679	1-235
680	1-26
682	1-58, 269-328
683	1-447
684	1-217
685	1-132
686	1-60
688	1-107
689	132-221, 327-377
690	1-388
691	1-141, 171-408
692	1-322
693	1-153
695	1-455
698	1-58, 117-174
699	240-300
700	1-159
701	1-69
702	1-175
703	1-298
704	1-136
705	1-168
706	1-419
707	1-382
708	8-245, 296-384
709	1-149
710	1-167
711	1-35
712	1-80, 116-156, 206-241
713	33-376
714	1-304
715	1-242
717	1-145
718	1-350
720	1-257
721	1-360
722	1-191
724	1-139
726	1-207
727	99-164
728	1-321
730	156-372
731	1-109, 256-290
735	25-192
737	1-160
738	1-227
739	441-514



742	217-280
743	10-275
747	1-179
749	2-31, 139-168
750	349-410
752	1-119
753	1-121
754	1-28
760	25-175
761	1-212
763	8-75
766	1-59, 102-248, 295-320
769	53-85
771	1-370
774	1-347
776	1-200
778	39-342
779	4-28
780	1-49, 407-472
781	116-426
782	1-59
783	1-53, 219-453
784	29-53, 219-263, 426-494
785	99-347, 386-461
786	2-28
788	1-279
789	1-58
790	226-268
792	129-218
794	265-431
796	5-86
797	1-34
799	1-344
802	46-477
806	64-384
807	135-301
808	2-314
810	6-39
1600	1-25
1601	1-225
1602	23-139
1603	1-294
1606	15-44
1607	1-361
1611	85-228
1612	1-221
1613	138-281
1614	65-171
1615	2-142
1616	1-46
1617	1-95
1620	1-187
1621	1-136

1622	32-280, 311-400
------	-----------------

5

Table IV

SEQ ID NO	Positions of Preferred Fragments
35	1-52
41	1-115
45	1-47
46	1-33
66	400-428
82	83-149
93	399-508
105	1-36
114	1-79
120	1-386
121	1-190
124	1-249
125	295-328
139	1-81, 125-268
159	1-139, 180-230
165	1-78
170	179-205, 248-280
194	1-150
213	1-158
247	1-104, 155-183, 280-359
269	31-143
350	139-386
368	228-446
385	1-72, 143-193
415	95-229
435	1-386
436	446-472
441	1-361
454	1-349
455	1-105
459	35-161, 200-311
460	1-26, 56-140
481	1-429
489	1-84
496	1-44, 84-215
501	153-430
502	1-91
504	1-63

505	1-68
514	1-303
515	237-351
519	1-145
526	231-366
530	1-88
535	1-55
570	76-207
576	168-218, 261-288
588	1-331
597	1-83
627	1-43
634	1-41
641	1-55, 334-483
672	1-34
687	1-129
708	1-245, 296-384
710	1-26, 104-167
722	1-191
730	1-465
731	1-43
735	1-91
737	1-160
738	1-186
739	1-48
742	1-62, 99-248
743	1-315, 412-459
744	1-31
747	1-63
749	1-32
750	1-38
752	1-139
753	1-193
754	1-28
759	1-38
760	1-115
763	1-62
765	1-126
769	1-85
770	1-40
771	1-148
774	1-134
775	265-531
776	71-203
777	333-469
778	144-468
779	1-28
780	1-49
781	1-102
782	1-59
783	1-53
784	1-220, 262-390
785	1-339, 408-461

786	1-28
789	1-58
791	1-126
792	1-31, 129-220
793	1-31
794	355-431
795	1-33
797	1-31
798	1-31
799	1-401
801	1-117
802	1-92
806	64-384
807	1-331
808	1-351
810	1-39
1600	1-25
1603	1-341
1606	1-31
1607	1-361
1608	164-305
1611	85-228
1612	1-221
1613	112-360
1614	1-171
1615	94-193
1617	1-155
1620	1-246

### III. Evaluation of Spatial and Temporal Expression of mRNAs Corresponding to the 5'ESTs, Consensus Contigated 5'ESTs, or EST-related nucleic acids

5

#### EXAMPLE 17

##### Expression Patterns of mRNAs From Which the 5'ESTs were obtained

Each of the SEQ ID NOs. 24-811 and 1600-1622 was also categorized based on the tissue from which its corresponding mRNA was obtained, as follows.

10 Table V shows the spatial distribution of each nucleic acid sequence of the invention (SEQ ID NOs. 24-811 and 1600-1622) referred to by its sequence identification number in the first column. In the second column entitled tissue distribution, the spatial distribution is represented by the number of individual 5'ESTs used to assemble the consensus contigated 5'ESTs for a given tissue. Each type of tissue listed in Table V is encoded by a letter. The correspondence between the letter code and the tissue  
15 type is given in Table VI.

Table V

SEQ ID NO	Tissue Distribution
24	AA:1
25	S:1
26	P:1
27	W:1
28	P:1
29	S:1
30	P:1
31	P:1
32	P:1
33	P:1
34	AB:1
35	G:3; P:1; S:1; W:3; AA:4
36	P:1
37	S:1
38	Q:1
39	P:1
40	AB:1
41	B:1; C:3; F:1; G:1; H:4; S:2; T:8; W:1; Z:1; AA:3; AC:1; AD:3
42	A:1
43	N:2
44	P:1
45	C:2; K:1; O:1; S:5
46	K:1; S:2; AA:1
47	AA:1
48	C:1; O:1; P:8
49	P:1
50	P:1
51	P:1
52	S:1
53	AA:1
54	T:1
55	P:1
56	P:1
57	P:1
58	P:1
59	P:7; T:2; Z:1
60	R:1
61	C:1
62	P:1
63	F:1
64	AA:1
65	F:1

66	P:4; T:2; Z:1
67	S:1
68	AA:1
69	P:1
70	P:1
71	S:1
72	W:1
73	G:1
74	P:1
75	N:1
76	P:1
77	S:1
78	U:1
79	B:1
80	P:1
81	AC:1
82	K:1; O:1
83	G:1
84	C:1; K:2; P:29; S:2; T:1; X:2; Y:1; AA:2
85	K:1
86	C:1
87	F:1
88	AB:1
89	H:1
90	M:1
91	B:1
92	K:1
93	AC:2
94	P:1
95	M:1
96	Z:2
97	K:1; P:11; S:1; X:1; AA:1
98	W:1
99	X:1
100	P:1
101	AB:1
102	F:1
103	AA:1
104	K:1
105	B:4; C:6; E:2; H:3; O:2; Q:1; S:3; AC:2
106	T:1
107	O:1
108	P:1
109	G:1
110	AA:1
111	T:1
112	P:1
113	F:1

114	B:3; C:4; K:5; S:4; Y:1
115	U:1
116	W:1
117	T:1
118	T:2
119	T:1
120	H:3
121	AA:3
122	K:1
123	H:2
124	AA:2
125	B:1; G:1; J:3; T:13; Y:5; AA:5; AD:2
126	H:1; P:1
127	K:1
128	F:1
129	G:1
130	P:1
131	B:1
132	AA:1
133	W:1
134	P:1
135	K:1
136	B:1; C:1
137	B:1
138	H:1
139	AC:2
140	T:1
141	B:1
142	H:1
143	T:1
144	H:1
145	B:1
146	R:1
147	P:1
148	C:1; H:2; O:1; S:2; T:1; AC:2
149	H:1
150	AA:1
151	W:1
152	S:1
153	F:1
154	M:1
155	B:1
156	R:1
157	W:1
158	T:1
159	C:1; AA:1
160	F:1
161	H:1

162	D:1
163	AA:1
164	AA:1
165	W:3
166	AA:1
167	W:1
168	F:1
169	B:1
170	G:2
171	E:1
172	B:1
173	F:1
174	B:1
175	W:1
176	K:1
177	AA:1
178	S:1
179	K:1
180	AA:1
181	W:1
182	K:1
183	T:1
184	P:1
185	B:1
186	W:1
187	R:1
188	T:1
189	T:1
190	W:1
191	A:1
192	F:1
193	B:1
194	G:3
195	W:1
196	O:1
197	T:1
198	O:1
199	B:1
200	AA:1
201	G:1
202	B:1
203	G:1
204	P:1
205	AA:1
206	Y:1
207	Y:1
208	AA:1
209	G:1



210	H:1
211	C:1
212	H:1
213	W:2
214	Y:1
215	AB:1
216	K:1
217	M:1
218	AD:1
219	A:1
220	AA:1
221	G:1
222	G:1
223	G:1; H:2; S:2; X:1
224	G:1
225	G:1
226	B:1
227	P:1
228	O:1
229	G:1
230	T:1
231	T:1
232	K:1
233	S:1
234	O:1
235	F:1
236	T:1
237	B:1
238	W:1
239	G:1
240	R:1
241	A:1
242	W:1
243	P:1
244	H:1
245	D:1
246	C:1
247	B:2
248	P:1
249	F:1
250	AB:1
251	W:1
252	H:1
253	B:1
254	S:1
255	T:1
256	W:1
257	T:1

258	AA:2
259	P:1
260	W:1
261	H:1
262	K:1
263	K:1
264	C:1; E:1; F:1; I:4; L:1; N:22; O:1; P:1; S:1; T:9; AA:1
265	A:1
266	T:1
267	K:1
268	H:1
269	T:2
270	T:1
271	T:1
272	B:1
273	Y:1
274	T:1
275	G:1
276	AA:1
277	T:1
278	AB:1
279	T:1
280	W:1
281	F:1
282	K:1
283	H:1
284	O:1
285	W:1
286	B:21; C:7; H:5; K:5; O:8; S:16; W:1; Y:3; Z:4; AA:2; AC:1
287	K:2; P:12; W:1; AC:2
288	S:1
289	K:2; P:8; W:1; AC:2
290	S:1
291	H:1
292	B:11; C:2; E:1; H:7; K:1; N:3; S:1; T:8; W:1; AA:28; AC:1
293	B:6; C:3; G:1; H:6; K:4; N:4; O:3; Q:2; S:5; T:1; U:1; V:2; Y:3; AA:1
294	B:1
295	H:1
296	AA:1
297	T:1
298	T:1
299	T:1
300	H:1; S:1
301	H:1
302	W:1
303	W:1
304	H:1
305	G:1

306	K:1
307	H:1
308	A:1
309	H:1
310	H:1
311	Y:1
312	G:1
313	H:1
314	K:1
315	Y:1
316	P:1
317	H:1
318	AA:1
319	H:1
320	O:1
321	Y:1
322	B:1
323	P:1
324	P:1
325	K:1
326	H:1
327	H:1
328	Q:1
329	S:1
330	B:1
331	T:1
332	T:1
333	B:1
334	T:1
335	W:1
336	P:1
337	A:1
338	AA:1
339	AA:1
340	G:1
341	C:1
342	K:1
343	S:1
344	G:1
345	B:1
346	Y:1
347	G:1
348	F:1
349	AA:5
350	B:15; C:1; G:1; H:1; O:1; Q:2; S:1; X:1; Y:1
351	F:1
352	R:1
353	O:1

354	H:1
355	W:1
356	F:1
357	T:1
358	S:1
359	X:1
360	T:1
361	K:1
362	K:1
363	G:1
364	K:1
365	G:1
366	AA:1
367	F:1
368	C:2; H:2; X:1
369	E:1
370	T:1
371	H:1
372	G:1
373	AA:1
374	G:1
375	F:1
376	F:1
377	R:1
378	AA:1
379	AA:1
380	C:1
381	H:1
382	T:1
383	W:1
384	S:1
385	AA:2
386	D:1
387	O:1
388	W:1
389	F:1
390	W:1
391	K:1
392	W:1
393	K:1
394	T:1
395	H:1
396	T:1
397	T:1
398	G:1
399	C:1
400	K:1
401	B:1

402	H:1
403	B:1
404	B:1
405	H:1
406	AB:1
407	O:1
408	P:1
409	X:1
410	H:1
411	B:9; C:3; K:3; L:2; O:1; S:2; X:1; AA:1
412	G:1; S:2; V:2; W:1; Y:1; Z:1
413	W:1
414	G:1
415	B:3; C:3; F:1; G:2; H:4; J:1; K:1; O:1; P:3; S:1; V:1
416	I:1
417	F:1
418	F:1
419	F:1
420	AA:1
421	F:1
422	T:1
423	P:1
424	B:1
425	Y:1
426	W:1
427	AA:1
428	W:1
429	H:1
430	Y:1
431	J:1
432	AA:1
433	G:1
434	AA:1
435	B:3; H:1
436	B:9; G:4; H:8; K:2; O:2; W:1; Z:2; AA:2; AD:3
437	H:1; T:1
438	T:1
439	R:1
440	M:1
441	H:2
442	W:1
443	B:1
444	W:1
445	AB:1
446	F:1
447	AD:1
448	AB:1
449	N:1

450	T:1
451	W:1
452	O:1
453	AA:1
454	D:28
455	W:1
456	T:1
457	G:1
458	W:1
459	Y:4
460	B:3
461	P:2
462	K:1
463	T:1
464	H:1
465	G:1
466	AC:1
467	R:1
468	S:1
469	B:1
470	S:1
471	T:1
472	AA:1
473	W:1
474	T:1
475	S:1
476	T:1
477	AA:1
478	G:1
479	W:1
480	B:1
481	O:2
482	K:1
483	P:1
484	W:1
485	P:1
486	B:1
487	Y:1
488	H:1
489	P:1; Q:1; S:3
490	C:1
491	S:1
492	H:1
493	B:1
494	H:1
495	G:1
496	N:2
497	B:1

498	G:1
499	P:1
500	G:1
501	C:1; K:1; Q:1
502	B:4
503	R:1
504	B:5; H:2; W:2
505	G:2; H:1
506	W:1
507	B:1
508	W:1
509	AB:1
510	H:1
511	N:1
512	J:1
513	AA:1
514	T:2
515	AA:5
516	F:1
517	C:1; O:1
518	W:1
519	T:4
520	B:1
521	H:1
522	H:2; T:3
523	H:1
524	AA:1
525	W:1
526	C:2; E:1; J:1; R:3; S:4; AA:1
527	H:1
528	S:1
529	P:1
530	B:1; H:1
531	O:1
532	Y:1
533	H:1
534	T:1
535	T:2
536	B:1
537	AD:1
538	AA:1
539	T:1
540	F:1
541	AD:1
542	W:1
543	W:1
544	F:1
545	T:1

546	F:1
547	K:1
548	Y:1
549	S:1
550	B:1
551	B:1
552	B:1
553	H:1
554	P:1
555	G:1
556	H:1
557	K:1
558	B:1
559	R:1
560	AB:1
561	C:1; S:1; V:1
562	AA:1
563	K:1
564	P:1
565	K:1
566	G:1
567	W:1
568	E:1; W:2
569	W:1
570	B:2
571	O:1
572	T:1
573	B:1
574	T:1
575	B:1
576	B:3
577	B:1
578	X:1
579	H:1
580	AA:1
581	AA:1
582	AA:1
583	AA:1
584	AA:1
585	D:1
586	H:1
587	H:1
588	AA:3
589	K:1
590	W:1
591	K:1
592	W:1
593	B:1



594	V:1
595	R:1
596	P:1
597	G:1; X:2; Z:1
598	X:1
599	F:1
600	F:1
601	Y:1
602	F:1
603	W:1
604	H:1
605	G:1
606	C:2; H:1; S:3; W:2; AD:3
607	W:1
608	C:1
609	F:1
610	K:1
611	M:1
612	AD:1
613	H:1
614	T:1
615	H:1
616	F:1
617	T:1
618	G:1
619	G:1
620	B:1
621	W:1
622	W:1
623	T:1
624	AA:1
625	G:1
626	M:1
627	C:2; T:2; W:1; Y:1
628	T:1
629	J:1
630	T:1
631	P:1
632	H:1
633	H:1
634	C:1; S:1; T:1; AD:1
635	J:1
636	G:1
637	W:1
638	AA:1
639	W:1
640	B:6; C:3; G:1; H:2; K:6; O:4; Q:1; R:2; S:1; T:3; Y:3; Z:2; AA:2; AC:2; AD:3

641	B:21; C:2; G:5; W:4; Y:1
642	AA:1
643	P:1
644	AA:1
645	T:1
646	K:1
647	F:1
648	F:1
649	F:1
650	T:1
651	W:1
652	T:1
653	T:1
654	P:1
655	B:1; H:2; N:1; T:3; Y:1
656	B:1
657	T:1
658	R:1
659	K:1
660	W:1
661	AA:1
662	Y:1
663	W:1
664	G:1
665	S:1
666	Y:1
667	F:1
668	T:1
669	B:1
670	F:1
671	T:1
672	A:2; B:6; C:1; G:1; H:3; J:1; L:1; P:2; Q:1; S:4; T:1; V:3; W:2; Y:1; AA:3; AD:2
673	T:1
674	G:1
675	F:1
676	M:1
677	G:1
678	Y:1
679	D:1
680	P:1
681	D:1
682	AA:1
683	G:1
684	K:1
685	G:1
686	P:1
687	B:3; C:2; D:2; E:2; J:4; V:2; AC:6

688	AA:1
689	S:1
690	AA:1
691	H:1
692	AA:1
693	S:1
694	AB:1
695	T:1
696	H:1
697	B:4; E:1; F:1; P:1; T:2; Z:2
698	O:1
699	W:1
700	S:1
701	O:1
702	B:1
703	AB:1
704	H:1
705	B:1
706	H:1
707	G:1
708	F:1; H:1; K:1; W:2; AA:1
709	H:1
710	T:2
711	C:1
712	G:1
713	Y:1
714	C:1
715	Y:1
716	Z:1
717	P:1
718	G:1
719	S:1
720	K:1
721	M:1
722	T:2
723	O:1; P:2; S:2
724	T:1
725	T:1
726	N:1
727	T:1
728	T:1
729	C:2; H:2; K:2; V:1; AC:1
730	B:7; H:2; Y:1
731	B:5; W:3
732	B:1; C:2; G:2; S:2; AA:9
733	B:6; C:2; G:1; H:10; O:2; P:6; Q:1; S:2; W:4; AC:2
734	B:6; O:1; V:1
735	C:1; O:2

736	B:1; H:2; N:1; T:3; Y:1
737	T:2
738	T:2
739	B:3; C:8; D:1; E:6; G:3; H:11; I:1; J:1; N:1; O:3; P:12; Q:3; S:2; T:2; W:1; AC:1; AD:8
740	H:2; Y:1
741	C:2; H:1
742	B:12; C:1; G:1; H:4; K:2; O:2; S:4; T:2; Y:2
743	AA:4
744	B:1; G:1; H:6; T:1; W:1
745	C:7; E:1; G:3; H:2; P:2; S:2; T:1; W:1; AD:2
746	G:2; S:1
747	T:2
748	S:3
749	H:1; O:2; S:2
750	Y:1; AD:1
751	B:8; G:2; H:2; I:1; Q:2; S:2; T:1; W:2
752	T:3
753	P:4
754	B:1; H:2
755	B:7; C:1; G:6; H:2; K:1; U:2; V:1; Z:1
756	C:1; H:1; J:2; O:2; S:1; T:2; W:1; AA:1
757	B:1; C:1; K:3; S:1; V:1; Y:1
758	E:1; H:2; K:1; P:1; Q:1; AD:5
759	B:6; C:1; Y:1
760	B:4
761	W:2
762	B:3; C:7; H:9; N:1; S:1; T:1; Y:1; AA:1
763	N:1; S:1; AA:5
764	H:3
765	B:3; G:1; W:1
766	H:2
767	C:1; AA:3
768	B:2; C:6; H:9; N:1; S:1; T:1; Y:1; AA:1
769	A:1; B:4; C:4; F:4; G:6; H:10; K:2; O:8; P:2; R:1; S:8; T:2; W:3; AA:2; AC:1
770	A:2; P:16; X:1
771	AA:3
772	O:4
773	B:1; C:1; W:1
774	P:2; X:4
775	B:18; C:6; H:5; K:3; O:7; S:10; W:1; Y:3; Z:2; AA:2; AC:1
776	H:7
777	B:26; C:8; H:5; K:4; O:10; S:17; W:1; Y:4; Z:4; AA:4; AC:2
778	B:6
779	B:3; C:1; G:1; H:2; K:1; Q:1; S:8; W:2; Y:9; AA:4
780	B:3; C:1; F:1; P:1; W:1; AC:1
781	I:2; N:1; P:1; R:3; AA:1
782	B:2

783	H:1; P:2; S:3; AD:1
784	H:1; P:1; S:4; AD:1
785	T:2
786	D:1; AC:9
787	H:1; L:1; S:1
788	B:6; S:4
789	S:1; T:1
790	B:1; C:2; H:5; W:1; AD:1
791	B:3; C:2; D:3; E:2; J:4; V:3; AC:5
792	B:3; D:1; K:2; S:2; Y:1
793	B:2; G:2; AA:1
794	B:25; C:4; D:1; E:1; F:3; G:6; J:1; K:6; N:1; O:1; P:2; R:1; S:3; T:2; W:2; X:1; Y:1; Z:1; AA:1; AC:2; AD:1
795	B:4; C:1; E:2; H:4; J:1; L:1; O:4; S:1; V:1; Y:3; Z:1
796	H:5
797	B:2; E:1; N:2
798	B:1; G:1; H:6; T:1; W:1
799	H:2
800	H:2; I:2; AA:1
801	A:2; B:4; C:14; D:1; H:2; K:1; N:2; S:4; T:1; W:2; AA:20
802	AA:17
803	B:2; G:3; H:3; S:1; U:1; AC:1; AD:2
804	C:1; S:2; T:2; X:2; AA:1; AC:1
805	B:5; C:6; D:5; H:17; J:2; K:4; N:1; O:6; P:2; S:5; T:5; W:1; X:1; Z:2; AA:13; AC:3
806	B:2; C:3; D:3; H:6; J:2; K:1; N:1; O:3; P:1; S:2; T:4; W:1; X:1; Z:1; AA:5; AC:1
807	H:1; AC:4
808	R:13
809	B:3; W:4
810	B:16; S:1; Y:14
811	B:8; C:5; G:1; H:1; K:5; O:2; Q:2; R:2; S:2; T:3; Y:4; Z:2; AA:1; AC:1; AD:2
1600	T:4
1601	AA:3
1602	C:3; H:1
1603	H:2; AC:2
1604	B:7; C:1; E:1; H:1; P:2; R:3; S:2; T:2; Z:3; AA:2
1605	C:4; H:3; O:1
1606	A:3; B:13; C:14; D:2; E:10; F:3; G:19; H:32; K:11; O:5; P:2; R:3; S:16; T:4; W:2; Y:10; Z:8; AA:1; AC:3
1607	T:3
1608	B:3; P:2
1609	R:4
1610	B:4
1611	B:3; T:1
1612	T:2
1613	V:5
1614	D:3

1615	AA:10
1616	B:4
1617	T:2
1618	K:2; S:8; AA:1
1619	B:2
1620	W:2
1621	H:1; AB:1
1622	H:2

Table VI

Tissue code	Tissue type
A	Bone Marrow
B	Brain
c	Cancerous prostate
D	Cerebellum
E	Colon
F	Dystrophic muscle
G	Fetal brain
H	Fetal kidney
I	Fetal liver
J	Heart
K	Hypertrophic prostate
L	Kidney
M	Large intestine
N	Liver
O	Lung
P	Lymph ganglia
Q	Lymphocytes
R	Muscle
S	Prostate
T	Ovary
U	Pancreas
V	Placenta
W	Spinal cord
X	Spleen
Y	Substantia nigra
Z	Surrenals
AA	Testis
AB	Thyroid
AC	Umbilical cord
AD	Uterus

- 5 In addition to categorizing the 5' ESTs and consensus contiguated 5' ESTs with respect to their tissue of origin, the spatial and temporal expression patterns of the mRNAs corresponding to the 5' ESTs and consensus contiguated 5' ESTs, as well as their expression levels, may be determined as described in Example 18 below.

Characterization of the spatial and temporal expression patterns and expression levels of these mRNAs is useful for constructing expression vectors capable of producing a desired level of gene product in a desired spatial or temporal manner, as will be discussed in more detail below.

Furthermore, 5' ESTs and consensus contigated 5' ESTs whose corresponding mRNAs are associated with disease states may also be identified. For example, a particular disease may result from the lack of expression, over expression, or under expression of a mRNA corresponding to a 5' EST or consensus contigated 5' EST. By comparing mRNA expression patterns and quantities in samples taken from healthy individuals with those from individuals suffering from a particular disease, 5' ESTs or consensus contigated 5' ESTs responsible for the disease may be identified.

It will be appreciated that the results of the above characterization procedures for 5' ESTs and consensus contigated 5' ESTs also apply to extended cDNAs (obtainable as described below) which contain sequences adjacent to the 5' ESTs and consensus contigated 5' ESTs. It will also be appreciated that if desired, characterization may be delayed until extended cDNAs have been obtained rather than characterizing the 5' ESTs or consensus contigated 5' ESTs themselves.

#### EXAMPLE 18

##### Evaluation of Expression Levels and Patterns of mRNAs

##### Corresponding to EST-Related Nucleic Acids

Expression levels and patterns of mRNAs corresponding to EST-related nucleic acids may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277. Briefly, an EST-related nucleic acid, fragment of an EST-related nucleic acid, positional segment of an EST-related nucleic acid, or fragment of a positional segment of an EST-related nucleic acid corresponding to the gene encoding the mRNA to be characterized is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the EST-related nucleic acid, fragment of an EST-related nucleic acid, positional segment of an EST-related nucleic acid, or fragment of a positional segment of an EST-related nucleic acid is 100 or more nucleotides in length. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (*i.e.* biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50°C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (*i.e.* RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

The EST-related nucleic acid, fragment of an EST-related nucleic acid, positional segment of an EST-related nucleic acid, or fragment of a positional segment of an EST-related nucleic acid may also be

tagged with nucleotide sequences for the serial analysis of gene expression (SAGE) as disclosed in UK Patent Application No. 2 305 241 A. In this method, cDNAs are prepared from a cell, tissue, organism or other source of nucleic acid for which gene expression patterns must be determined. The resulting cDNAs are separated into two pools. The cDNAs in each pool are cleaved with a first restriction  
5 endonuclease, called an anchoring enzyme, having a recognition site which is likely to be present at least once in most cDNAs. The fragments which contain the 5' or 3' most region of the cleaved cDNA are isolated by binding to a capture medium such as streptavidin coated beads. A first oligonucleotide linker having a first sequence for hybridization of an amplification primer and an internal restriction site for a so called tagging endonuclease is ligated to the digested cDNAs in the first pool. Digestion with the  
10 second endonuclease produces short tag fragments from the cDNAs.

A second oligonucleotide having a second sequence for hybridization of an amplification primer and an internal restriction site is ligated to the digested cDNAs in the second pool. The cDNA fragments in the second pool are also digested with the tagging endonuclease to generate short tag fragments derived from the cDNAs in the second pool. The tags resulting from digestion of the first and second  
15 pools with the anchoring enzyme and the tagging endonuclease are ligated to one another to produce so called ditags. In some embodiments, the ditags are concatamerized to produce ligation products containing from 2 to 200 ditags. The tag sequences are then determined and compared to the sequences of the EST-related nucleic acid, fragment of an EST-related nucleic acid, positional segment of an EST-related nucleic acid, or fragment of a positional segment of an EST-related nucleic acid to determine  
20 which 5' ESTs, consensus contigated 5' ESTs, or extended cDNAs are expressed in the cell, tissue, organism, or other source of nucleic acids from which the tags were derived. In this way, the expression pattern of the 5' ESTs, consensus contigated 5' ESTs, or extended cDNAs in the cell, tissue, organism, or other source of nucleic acids is obtained.

Quantitative analysis of gene expression may also be performed using arrays. As used herein,  
25 the term array means a one dimensional, two dimensional, or multidimensional arrangement of EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids. Preferably, the EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids are at least 10, 12, 15, 18, 20, 23, 25, 28,  
30 30, 35, 40, or 50 nucleotides in length. More preferably, the EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids are at least 100 nucleotide long. More preferably, the fragments are more than 100 nucleotides in length. In some embodiments, the EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments EST-related nucleic acids, or fragments of  
35 positional segments of EST-related nucleic acids may be more than 500 nucleotides long.

For example, quantitative analysis of gene expression may be performed with EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments EST-related nucleic acids, or



fragments of positional segments of EST-related nucleic acids in a complementary DNA microarray as described by Schena *et al.* (*Science* 270:467-470, 1995; *Proc. Natl. Acad. Sci. U.S.A.* 93:10614-10619, 1996). EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids are amplified by  
5 PCR and arrayed from 96-well microtiter plates onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C.

- 10 Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm<sup>2</sup> microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1 x SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1 x SSC/0.2% SDS). Arrays are scanned in 0.1 x SSC using a fluorescence laser scanning device fitted with a custom  
15 filter set. Accurate differential expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of the expression of genes may also be performed with EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids in complementary DNA arrays as  
20 described by Pietu *et al.* (*Genome Research* 6:492-503, 1996). The EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids thereof are PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by phospho-  
25 imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis of the EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids can be done through high density nucleotide arrays as described by Lockhart  
30 *et al.* (*Nature Biotechnology* 14: 1675-1680, 1996) and Sosnowsky *et al.* (*Proc. Natl. Acad. Sci.* 94:1119-1123, 1997). Oligonucleotides of 15-50 nucleotides corresponding to sequences of EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids are synthesized directly on the chip (Lockhart *et al.*, *supra*) or synthesized and then addressed to the chip (Sosnowsky *et al.*, *supra*).  
35 Preferably, the oligonucleotides are about 20 to 25 nucleotides in length.

cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an

average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart *et al, supra* and application of different electric fields (Sonowsky *et al, supra.*), the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target  
5 oligonucleotide in different cDNA samples indicates a differential expression of the mRNA corresponding to the 5' EST, consensus contigated 5' EST or extended cDNA from which the oligonucleotide sequence has been designed.

#### IV. Use of 5' ESTs to Clone Extended cDNAs and to Clone the Corresponding Genomic DNAs

10 Once 5' ESTs or consensus contigated 5' ESTs which include the 5' end of the corresponding mRNAs have been selected using the procedures described above, they can be utilized to isolate extended cDNAs which contain sequences adjacent to the 5' ESTs or consensus contigated 5' ESTs. The extended cDNAs may include the entire coding sequence of the protein encoded by the corresponding mRNA, including the authentic translation start site. If the extended cDNA encodes a  
15 secreted protein, it may contain the signal sequence, and the sequence encoding the mature protein remaining after cleavage of the signal peptide.

Extended cDNAs which include the entire coding sequence of the protein encoded by the corresponding mRNA are referred to herein as "full-length cDNAs." Alternatively, the extended cDNAs may not include the entire coding sequence of the protein encoded by the corresponding mRNA,  
20 although they do include sequences adjacent to the 5'ESTs or consensus contigated 5' ESTs. In some embodiments in which the extended cDNAs are derived from an mRNA encoding a secreted protein, the extended cDNAs may include only the sequence encoding the mature protein remaining after cleavage of the signal peptide, or only the sequence encoding the signal peptide.

Examples 19 and 20 below describe a general method for obtaining extended cDNAs using 5'  
25 ESTs or consensus contigated 5' ESTs and nucleic acid homologous thereto. Example 21 below describes the cloning and sequencing of several extended cDNAs, including full-length cDNAs which include the authentic 5' end of the corresponding mRNA for several secreted proteins.

The methods of Examples 19 and 20 can also be used to obtain extended cDNAs which encode less than the entire coding sequence of proteins encoded by the genes corresponding to the 5' ESTs or  
30 consensus contigated 5'ESTs. In some embodiments, the extended cDNAs isolated using these methods encode at least 5,10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids of one of the proteins encoded by the sequences of SEQ ID NOs. 24-811 and 1600-1622. In some embodiments, the extended cDNAs isolated using these methods encode at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100,  
35 811. or 150 consecutive amino acids of one of the proteins encoded by the sequences of SEQ ID NOs. 24-

#### EXAMPLE 19

General Method for Using 5' ESTs or Consensus Contigated 5'ESTs to Clone and Sequence Extended cDNAs which Include the Entire Coding Region and the Authentic 5'End of the Corresponding mRNA

The following general method may be used to quickly and efficiently isolate extended cDNAs including sequence adjacent to the sequences of the 5' ESTs or Consensus Contigated 5'ESTs used to obtain them. This method may be applied to obtain extended cDNAs for any 5' EST or consensus contigated 5' EST of the invention, including those 5' ESTs and consensus contigated 5' ESTs encoding secreted proteins. This method is illustrated in Figure 3.

1. Obtaining Extended cDNAs

The method takes advantage of the known 5' sequence of the mRNA. A reverse transcription reaction is conducted on purified mRNA with a poly dT primer containing a nucleotide sequence at its 5' end allowing the addition of a known sequence at the end of the cDNA which corresponds to the 3' end of the mRNA. Such a primer and a commercially-available reverse transcriptase enzyme are added to a buffered mRNA sample yielding a reverse transcript anchored at the 3' polyA site of the RNAs. Nucleotide monomers are then added to complete the first strand synthesis.

After removal of the mRNA hybridized to the first cDNA strand by alkaline hydrolysis, the products of the alkaline hydrolysis and the residual poly dT primer can be eliminated with an exclusion column.

Subsequently, a pair of nested primers on each end is designed based on the known 5' sequence from the 5' EST or consensus contigated 5' EST and the known 3' end added by the poly dT primer used in the first strand synthesis. Software used to design primers are either based on GC content and melting temperatures of oligonucleotides, such as OSP (Illier and Green, *PCR Meth. Appl.* 1:124-128, 1991), or based on the octamer frequency disparity method (Griffais *et al.*, *Nucleic Acids Res.* 19: 3887-3891, 1991 ) such as PC-Rare ([http:// bioinformatics.weizmann.ac.il/software/PC-Rare/doc/manuel.html](http://bioinformatics.weizmann.ac.il/software/PC-Rare/doc/manuel.html)). Preferably, the nested primers at the 5' end and the nested primers at the 3' end are separated from one another by four to nine bases. These primer sequences may be selected to have melting temperatures and specificities suitable for use in PCR.

A first PCR run is performed using the outer primer from each of the nested pairs. A second PCR run using the inner primer from each of the nested pairs is then performed on a small sample of the first PCR product. Thereafter, the primers and remaining nucleotide monomers are removed.

2. Sequencing Extended cDNAs or Fragments Thereof

Due to the lack of position constraints on the design of 5' nested primers compatible for PCR use using the OSP software, amplicons of two types are obtained. Preferably, the second 5' primer is located upstream of the translation initiation codon thus yielding a nested PCR product containing the entire coding sequence. Such an extended cDNA may be used in a direct cloning procedure as described in section a below. However, in some cases, the second 5' primer is located downstream of the translation initiation codon, thereby yielding a PCR product containing only part of the ORF. Such incomplete PCR products are submitted to a modified procedure described in section b below.

a) *Nested PCR products containing complete ORFs*

When the resulting nested PCR product contains the complete coding sequence, as predicted from the 5' EST or consensus contigated 5' EST sequence, it is directly cloned in an appropriate vector as described in section 3.

5 b) *Nested PCR products containing incomplete ORFs*

When the amplicon does not contain the complete coding sequence, intermediate steps are necessary to obtain both the complete coding sequence and a PCR product containing the full coding sequence. The complete coding sequence can be assembled from several partial sequences determined directly from different PCR products.

10 Once the full coding sequence has been completely determined, new primers compatible for PCR use are then designed to obtain amplicons containing the whole coding region. However, in such cases, 3' primers compatible for PCR use are located inside the 3' UTR of the corresponding mRNA, thus yielding amplicons which lack part of this region, *i.e.* the polyA tract and sometimes the polyadenylation signal, as illustrated in Figure 3. Such extended cDNAs are then cloned into an  
15 appropriate vector as described in section 3.

c) *Sequencing extended cDNAs*

Sequencing of extended cDNAs can be performed using a Die Terminator approach with the AmpliTaq DNA polymerase FS kit available from Perkin Elmer.

In order to sequence long PCR fragments, primer walking is performed using software such as  
20 OSP to choose primers and automated computer software such as ASMG (Sutton *et al.*, *Genome Science Technol.* 1: 9-19, 1995) to construct contigs of walking sequences including the initial 5' tag. Preferably, primer walking is performed until the sequences of full length cDNAs are obtained.

Completion of the sequencing of a given extended cDNA fragment may be assessed by comparing the sequence length to the size of the corresponding nested PCR product. When Northern  
25 blot data are available, the size of the mRNA detected for a given PCR product may also be used to finally assess that the sequence is complete. Sequences which do not fulfill these criteria are discarded and will undergo a new isolation procedure.

3. Cloning Extended cDNAs

The PCR product containing the full coding sequence is then cloned in an appropriate vector.  
30 For example, the extended cDNAs can be cloned into any expression vector known in the art, such as pED6dpc2 (DiscoverEase, Genetics Institute, Cambridge, MA).

Cloned PCR products are then entirely sequenced in order to obtain at least two sequences per clone. Preferably, the sequences are obtained from both sense and antisense strands according to the aforementioned procedure with the following modifications. First, both 5' and 3' ends of cloned  
35 PCR products are sequenced in order to confirm the identity of the clone. Second, primer walking is performed if the full coding region has not been obtained yet. Contiguation is then performed using primer walking sequences for cloned products as well as walking sequences that have already

contigated for uncloned PCR products. The sequence is considered complete when the resulting contigs include the whole coding region as well as overlapping sequences with vector DNA on both ends. All the contigated sequences for each cloned amplicon are then used to obtain a consensus sequence.

5 4. Selection of Cloned Full length Sequences

*a) Computer analysis of extended cDNAs*

Following identification of contaminants and masking of repeats, structural features, e.g. polyA tail and polyadenylation signal, of the sequences of extended cDNAs are subsequently determined using methods known to those skilled in the art. For example, algorithm, parameters and  
10 criteria defined in Figure 10 may be used. Briefly, a polyA tail is defined as a homopolymeric stretch of at least 11 A with at most one alternative base within it. The polyA tail search is restricted to the last 20 nucleotides of the sequence and limited to stretches of 11 consecutive A's because sequencing reactions are often not readable after such a polyA stretch. To search for a polyadenylation signal, the polyA tail is clipped from the full-length sequence. The 50 nucleotides preceding the polyA tail  
15 are searched for the canonic polyadenylation AAUAAA signal allowing one mismatch to account for possible sequencing errors as well as known variation in the canonical sequence of the polyadenylation signal.

Functional features, e.g. ORFs and signal sequences, of the sequences of extended cDNAs are subsequently determined as follows. The 3 upper strand frames of extended cDNAs are searched for  
20 ORFs defined as the maximum length fragments beginning with a translation initiation codon and ending with a stop codon. ORFs encoding at least 80 amino acids are preferred. If extended cDNAs encoding secreted proteins are desired, each found ORF is then scanned for the presence of a signal peptide using the matrix method described in Example 13.

Sequences of extended cDNAs are then compared, on a nucleotidic or proteic basis, to public  
25 sequences available at the time of filing.

*b) Selection of full-length cDNAs of interest*

A negative selection may then be performed in order to eliminate unwanted cloned sequences resulting from either contaminants or PCR artifacts as follows. Sequences matching contaminant sequences such as vector DNA, tRNA, mtRNA, rRNA sequences are discarded as well as those  
30 encoding ORF sequences exhibiting extensive homology to repeats. Sequences obtained by direct cloning (section 1a) but lacking polyA tail may be discarded. Only ORFs ending either before the polyA tail (section 1a) or before the end of the cloned 3'UTR (section 1b) may be selected. If extended cDNAs encoding secreted proteins are desired, ORFs containing a signal peptide are considered. In addition, ORFs containing unlikely mature proteins such as mature proteins which size is less than 20 amino acids  
35 or less than 25% of the immature protein size may be eliminated.

Then, for each remaining full length cDNA containing several ORFs, a preselection of ORFs may be performed using the following criteria. The longest ORF is preferred. If extended cDNAs

encoding secreted proteins are desired and if the ORF sizes are similar, the chosen ORF is the one which signal peptide has the highest score according to Von Heijne method.

Sequences of full length cDNA clones may then be compared pairwise after masking of the repeat sequences. Full-length cDNA sequences exhibiting extensive homology may be clustered in the same class. Each cluster may then be subjected to a cluster analysis that detects sequences resulting from internal priming or from alternative splicing, identical sequences or sequences with several frameshifts. A selection may be operated between clones belonging to the same class in order to detect clones encoding homologous but distinct ORFs which may be both selected if they both contain sequences of interest.

10 Selection of full-length cDNA clones encoding sequences of interest may subsequently be performed using the following criteria. Structural parameters (initial tag, polyadenylation site and signal) are first checked. Then, homologies with known nucleic acids and proteins are examined in order to determine whether the clone sequence match a known nucleotide/protein sequence and, in the latter case, its covering rate and the date at which the sequence became public. If there is no extensive  
15 match with sequences other than ESTs or genomic DNA, or if the clone sequence brings substantial new information, such as encoding a protein resulting from alternative splicing of an mRNA coding for an already known protein, the sequence is kept. Examples of such cloned full-length cDNAs containing sequences of interest are described in Example 21. Sequences resulting from chimera or double inserts or located on chromosome breaking points as assessed by homology to other sequences may be  
20 discarded during this procedure.

Extended cDNAs prepared as described above may be subsequently engineered to obtain nucleic acids which include desired portions of the extended cDNA using conventional techniques such as subcloning, PCR, or *in vitro* oligonucleotide synthesis. For example, nucleic acids which include only the full coding sequences may be obtained using techniques known to those skilled in the art.  
25 Alternatively, conventional techniques may be applied to obtain nucleic acids which contain only part of the coding sequences. In the case of nucleic acids encoding secreted proteins, nucleic acids containing only the coding sequence for the mature protein remaining after the signal peptide is cleaved off or nucleic acids which contain only the coding sequences for the signal peptides may be obtained.

Similarly, nucleic acids containing any other desired portion of the coding sequences for the  
30 encoded protein may be obtained. For example, the nucleic acid may contain at least 10, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 or 500 consecutive bases of an extended cDNA.

Once an extended cDNA has been obtained, it can be sequenced to determine the amino acid sequence it encodes. Once the encoded amino acid sequence has been determined, one can create and identify any of the many conceivable cDNAs that will encode that protein by simply using the  
35 degeneracy of the genetic code. For example, allelic variants or other homologous nucleic acids can be identified as described below. Alternatively, nucleic acids encoding the desired amino acid sequence can be synthesized *in vitro*.

In a preferred embodiment, the coding sequence may be selected using the known codon or codon pair preferences for the host organism in which the cDNA is to be expressed.

In addition to PCR based methods for obtaining cDNAs which include the authentic 5' end of the corresponding mRNA as well as the complete protein coding sequence of the corresponding mRNA, traditional hybridization based methods may also be employed. These methods may also be used to obtain the genomic DNAs which encode the mRNAs from which the 5' ESTs or consensus contigated 5' ESTS were derived, mRNAs corresponding to the extended cDNAs, or nucleic acids which are homologous to extended cDNAs, 5' ESTs, or consensus contigated 5' ESTs. Example 19 below provides examples of such methods.

#### EXAMPLE 20

Methods for Obtaining Extended cDNAs which Include the Entire Coding Region and the Authentic 5' End of the Corresponding mRNA or Nucleic Acids Homologous to Extended cDNAs, 5' ESTs or Consensus Contigated 5' ESTs

A full-length cDNA library can be made using the strategies described in Example 7. Alternatively, a cDNA library or genomic DNA library may be obtained from a commercial source or made using techniques familiar to those skilled in the art.

Such cDNA or genomic DNA libraries may be used to isolate extended cDNAs obtained from 5' ESTs or consensus contigated 5' ESTs or nucleic acids homologous to extended cDNAs, 5' ESTs, or consensus contigated 5' ESTs as follows. The cDNA library or genomic DNA library is hybridized to a detectable probe. The detectable probe may comprise at least 10, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 or 500 consecutive nucleotides of the 5' EST, consensus contigated 5' EST, or extended cDNA.

Techniques for identifying cDNA clones in a cDNA library which hybridize to a given probe sequence are disclosed in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual 2d Ed.*, Cold Spring Harbor Laboratory Press, 1989. The same techniques may be used to isolate genomic DNAs. Briefly, cDNA or genomic DNA clones which hybridize to the detectable probe are identified and isolated for further manipulation as follows. The detectable probe described in the preceding paragraph is labeled with a detectable label such as a radioisotope or a fluorescent molecule. Techniques for labeling the probe are well known and include phosphorylation with polynucleotide kinase, nick translation, *in vitro* transcription, and non radioactive techniques. The cDNAs or genomic DNAs in the library are transferred to a nitrocellulose or nylon filter and denatured. After blocking of non specific sites, the filter is incubated with the labeled probe for an amount of time sufficient to allow binding of the probe to cDNAs or genomic DNAs containing a sequence capable of hybridizing thereto.

By varying the stringency of the hybridization conditions used to identify cDNAs or genomic DNAs which hybridize to the detectable probe, cDNAs or genomic DNAs having different levels of homology to the probe can be identified and isolated as described below.

### 1. Identification of cDNA or Genomic DNA Sequences Having a High Degree of Homology to the Labeled Probe

To identify cDNAs or genomic DNAs having a high degree of homology to the probe sequence, the melting temperature of the probe may be calculated using the following formulas:

- 5 For probes between 14 and 70 nucleotides in length the melting temperature ( $T_m$ ) is calculated using the formula:  $T_m = 81.5 + 16.6(\log(Na^+)) + 0.41(\text{fraction G+C}) - (600/N)$  where N is the length of the probe.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation  $T_m = 81.5 + 16.6(\log(Na^+)) + 0.41(\text{fraction G+C}) - (0.63\%$

- 10 formamide) - (600/N) where N is the length of the probe.

Prehybridization may be carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100  $\mu$ g denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100  $\mu$ g denatured fragmented salmon sperm DNA, 50% formamide. The formulas for SSC and Denhardt's solutions are listed in Sambrook *et al.*, *supra*.

- 15 Hybridization is conducted by adding the detectable probe to the prehybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured before addition to the hybridization solution. The filter is contacted with the hybridization solution for a sufficient period of time to allow the probe to hybridize to extended cDNAs or genomic DNAs containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the
- 20 hybridization may be carried out at 15-25°C below the  $T_m$ . For shorter probes, such as oligonucleotide probes, the hybridization may be conducted at 15-25°C below the  $T_m$ . Preferably, for hybridizations in 6X SSC, the hybridization is conducted at approximately 68°C. Preferably, for hybridizations in 50% formamide containing solutions, the hybridization is conducted at approximately 42°C.

All of the foregoing hybridizations would be considered to be under "stringent" conditions.

- 25 Following hybridization, the filter is washed in 2X SSC, 0.1% SDS at room temperature for 15 minutes. The filter is then washed with 0.1X SSC, 0.5% SDS at room temperature for 30 minutes to 1 hour. Thereafter, the solution is washed at the hybridization temperature in 0.1X SSC, 0.5% SDS. A final wash is conducted in 0.1X SSC at room temperature.

- cDNAs or genomic DNAs which have hybridized to the probe are identified by autoradiography
- 30 or other conventional techniques.

### 2. Obtaining cDNA or Genomic DNA Sequences Having Lower Degrees of Homology to the Labeled Probe

The above procedure may be modified to identify cDNAs or genomic DNAs having decreasing levels of homology to the probe sequence. For example, to obtain cDNAs or genomic DNAs of

- 35 decreasing homology to the detectable probe, less stringent conditions may be used. For example, the hybridization temperature may be decreased in increments of 5°C from 68°C to 42°C in a hybridization buffer having a sodium concentration of approximately 1M. Following hybridization, the filter may be



washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate" conditions above 50°C and "low" conditions below 50°C.

Alternatively, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42°C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50°C. These conditions are considered to be "moderate" conditions above 25% formamide and "low" conditions below 25% formamide. cDNAs or genomic DNAs which have hybridized to the probe are identified by autoradiography.

10 3. Determination of the Degree of Homology between the Obtained cDNAs or Genomic DNAs and 5'ESTs, Consensus Contigated 5'ESTs, or Extended cDNAs or Between the Polypeptides Encoded by the Obtained cDNAs or Genomic DNAs and the Polypeptides Encoded by the 5'ESTs, Consensus Contigated 5'ESTs, or Extended cDNAs

To determine the level of homology between the hybridized cDNA or genomic DNA and the 5'EST, consensus contigated 5'EST or extended cDNA from which the probe was derived, the nucleotide sequences of the hybridized nucleic acid and the 5'EST, consensus contigated 5'EST or extended cDNA from which the probe was derived are compared. The sequences of the 5'EST, consensus contigated 5'EST or extended cDNA from which the probe was derived and the sequences of the cDNA or genomic DNA which hybridized to the detectable probe may be stored on a computer readable medium as described below and compared to one another using any of a variety of algorithms familiar to those skilled in the art, those described below.

To determine the level of homology between the polypeptide encoded by the hybridizing cDNA or genomic DNA and the polypeptide encoded by the 5'EST, consensus contigated 5'EST or extended cDNA from which the probe was derived, the polypeptide sequence encoded by the hybridized nucleic acid and the polypeptide sequence encoded by the 5'EST, consensus contigated 5'EST or extended cDNA from which the probe was derived are compared. The sequences of the polypeptide encoded by the 5'EST, consensus contigated 5'EST or extended cDNA from which the probe was derived and the polypeptide sequence encoded by the cDNA or genomic DNA which hybridized to the detectable probe may be stored on a computer readable medium as described below and compared to one another using any of a variety of algorithms familiar to those skilled in the art, those described below.

Protein and/or nucleic acid sequence homologies may be evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988, *Proc. Natl. Acad. Sci. USA* 85(8):2444-2448; Altschul *et al.*, 1990, *J. Mol. Biol.* 215(3):403-410; Thompson *et al.*, 1994, *Nucleic Acids Res.* 22(2):4673-4680; Higgins *et al.*, 1996, *Methods Enzymol.* 266:383-402; Altschul *et al.*, 1990, *J. Mol. Biol.* 215(3):403-410; Altschul *et al.*, 1993, *Nature Genetics* 3:266-272).

In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well known in the art (see, e.g., Karlin and Altschul, 1990, *Proc. Natl. Acad. Sci. USA* 87:2267-2268; Altschul *et al.*, 1990, *J. Mol. Biol.* 215:403-410; Altschul *et al.*, 1993, *Nature Genetics* 3:266-272; Altschul *et al.*, 1997, *Nuc. Acids Res.* 25:3389-3402). In particular, five specific BLAST programs are used to perform the following task:

- (1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;
- (2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;
- (3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;
- (4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and
- (5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (*i.e.*, aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet *et al.*, 1992, *Science* 256:1443-1445; Henikoff and Henikoff, 1993, *Proteins* 17:49-61). Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978, *Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure*, Washington: National Biomedical Research Foundation)

The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, e.g., Karlin and Altschul, 1990, *Proc. Natl. Acad. Sci. USA* 87:2267-2268).

The parameters used with the above algorithms may be adapted depending on the sequence length and degree of homology studied. In some embodiments, the parameters may be the default parameters used by the algorithms in the absence of instructions from the user.

In some embodiments, the level of homology between the hybridized nucleic acid and the extended cDNA, 5'EST, or 5' consensus contigated 5'EST from which the probe was derived may be determined using the FASTDB algorithm described in Brutlag *et al.* *Comp. App. Biosci.* 6:237-245, 1990. In such analyses the parameters may be selected as follows: Matrix=Unitary, k-tuple=4,

Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the sequence which hybridizes to the probe, whichever is shorter. Because the FASTDB program does not consider 5' or 3' truncations when calculating homology levels, if the sequence which hybridizes to the probe is truncated relative to the sequence of the extended cDNA, 5'EST, or consensus contigated 5'EST from which the probe was derived the homology level is manually adjusted by calculating the number of nucleotides of the extended cDNA, 5'EST, or consensus contigated 5' EST which are not matched or aligned with the hybridizing sequence, determining the percentage of total nucleotides of the hybridizing sequence which the non-matched or non-aligned nucleotides represent, and subtracting this percentage from the homology level. For example, if the hybridizing sequence is 700 nucleotides in length and the extended cDNA, 5'EST, or consensus contigated 5' EST sequence is 1000 nucleotides in length wherein the first 300 bases at the 5' end of the extended cDNA, 5'EST, or consensus contigated 5' EST are absent from the hybridizing sequence, and wherein the overlapping 700 nucleotides are identical, the homology level would be adjusted as follows. The non-matched, non-aligned 300 bases represent 30% of the length of the extended cDNA, 5'EST, or consensus contigated 5' EST. If the overlapping 700 nucleotides are 100% identical, the adjusted homology level would be  $100-30=70\%$  homology. It should be noted that the preceding adjustments are only made when the non-matched or non-aligned nucleotides are at the 5' or 3' ends. No adjustments are made if the non-matched or non-aligned sequences are internal or under any other conditions.

For example, using the above methods, nucleic acids having at least 95% nucleic acid homology, at least 96% nucleic acid homology, at least 97% nucleic acid homology, at least 98% nucleic acid homology, at least 99% nucleic acid homology, or more than 99% nucleic acid homology to the extended cDNA, 5'EST, or consensus contigated 5' EST from which the probe was derived may be obtained and identified. Such nucleic acids may be allelic variants or related nucleic acids from other species. Similarly, by using progressively less stringent hybridization conditions one can obtain and identify nucleic acids having at least 90%, at least 85%, at least 80% or at least 75% homology to the extended cDNA, 5'EST, or consensus contigated 5' EST from which the probe was derived.

Using the above methods and algorithms such as FASTA with parameters depending on the sequence length and degree of homology studied, for example the default parameters used by the algorithms in the absence of instructions from the user, one can obtain nucleic acids encoding proteins having at least 99%, at least 98%, at least 97%, at least 96%, at least 95%, at least 90%, at least 85%, at least 80% or at least 75% homology to the protein encoded by the extended cDNA, 5'EST, or consensus contigated 5' EST from which the probe was derived. In some embodiments, the homology levels can be determined using the "default" opening penalty and the "default" gap penalty, and a scoring matrix such as PAM 250 (a standard scoring matrix; see Dayhoff *et al.*, in: Atlas of Protein Sequence and Structure, Vol. 5, Supp. 3 (1978)).

Alternatively, the level of polypeptide homology may be determined using the FASTDB algorithm described by Brutlag *et al.* Comp. App. Biosci. 6:237-245, 1990. In such analyses the parameters may be selected as follows: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=Sequence Length, Gap  
5 Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the homologous sequence, whichever is shorter. If the homologous amino acid sequence is shorter than the amino acid sequence encoded by the extended cDNA, 5'EST, or consensus contigated 5' EST as a result of an N terminal and/or C terminal deletion the results may be manually corrected as follows. First, the number of amino acid residues of the amino acid sequence encoded by the extended cDNA, 5'EST, or consensus  
10 contigated 5' EST which are not matched or aligned with the homologous sequence is determined. Then, the percentage of the length of the sequence encoded by the extended cDNA, 5'EST, or consensus contigated 5' EST which the non-matched or non-aligned amino acids represent is calculated. This percentage is subtracted from the homology level. For example wherein the amino acid sequence encoded by the extended cDNA, 5'EST, or consensus contigated 5' EST is 100 amino acids in length  
15 and the length of the homologous sequence is 80 amino acids and wherein the amino acid sequence encoded by the extended cDNA or 5'EST is truncated at the N terminal end with respect to the homologous sequence, the homology level is calculated as follows. In the preceding scenario there are 20 non-matched, non-aligned amino acids in the sequence encoded by the extended cDNA, 5'EST, or consensus contigated 5' EST. This represents 20% of the length of the amino acid sequence encoded by  
20 the extended cDNA, 5'EST, or consensus contigated 5' EST. If the remaining amino acids are 100% identical between the two sequences, the homology level would be  $100\% - 20\% = 80\%$  homology. No adjustments are made if the non-matched or non-aligned sequences are internal or under any other conditions.

In addition to the above described methods, other protocols are available to obtain extended  
25 cDNAs using 5' ESTs or consensus contigated 5'ESTs as outlined in the following paragraphs.

Extended cDNAs may be prepared by obtaining mRNA from the tissue, cell, or organism of interest using mRNA preparation procedures utilizing polyA selection procedures or other techniques known to those skilled in the art. A first primer capable of hybridizing to the polyA tail of the mRNA is hybridized to the mRNA and a reverse transcription reaction is performed to generate a first cDNA  
30 strand.

The first cDNA strand is hybridized to a second primer containing at least 10 consecutive nucleotides of the sequences of SEQ ID NOs 24-811 and 1600-1622. Preferably, the primer comprises at least 10, 12, 15, 17, 18, 20, 23, 25, or 28 consecutive nucleotides from the sequences of SEQ ID NOs 24-811 and 1600-1622. In some embodiments, the primer comprises more than 30 nucleotides from the  
35 sequences of SEQ ID NOs 24-811 and 1600-1622. If it is desired to obtain extended cDNAs containing the full protein coding sequence, including the authentic translation initiation site, the second primer used contains sequences located upstream of the translation initiation site. The second primer is

extended to generate a second cDNA strand complementary to the first cDNA strand. Alternatively, RT-PCR may be performed as described above using primers from both ends of the cDNA to be obtained.

Extended cDNAs containing 5' fragments of the mRNA may be prepared by hybridizing an mRNA comprising the sequences of SEQ ID NOs. 24-811 and 1600-1622 with a primer comprising a complementary to a fragment of an EST-related nucleic acid hybridizing the primer to the mRNAs, and reverse transcribing the hybridized primer to make a first cDNA strand from the mRNAs. Preferably, the primer comprises at least 10, 12, 15, 17, 18, 20, 23, 25, or 28 consecutive nucleotides of the sequences complementary to SEQ ID NOs. 24-811 and 1600-1622.

Thereafter, a second cDNA strand complementary to the first cDNA strand is synthesized. The second cDNA strand may be made by hybridizing a primer complementary to sequences in the first cDNA strand to the first cDNA strand and extending the primer to generate the second cDNA strand.

The double stranded extended cDNAs made using the methods described above are isolated and cloned. The extended cDNAs may be cloned into vectors such as plasmids or viral vectors capable of replicating in an appropriate host cell. For example, the host cell may be a bacterial, mammalian, avian, or insect cell.

Techniques for isolating mRNA, reverse transcribing a primer hybridized to mRNA to generate a first cDNA strand, extending a primer to make a second cDNA strand complementary to the first cDNA strand, isolating the double stranded cDNA and cloning the double stranded cDNA are well known to those skilled in the art and are described in *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. 1997 and Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989.

Alternatively, other procedures may be used for obtaining full-length cDNAs or extended cDNAs. In one approach, full-length or extended cDNAs are prepared from mRNA and cloned into double stranded phagemids as follows. The cDNA library in the double stranded phagemids is then rendered single stranded by treatment with an endonuclease, such as the Gene II product of the phage F1 and an exonuclease (Chang *et al.*, *Gene* 127:95-8, 1993). A biotinylated oligonucleotide comprising the sequence of a fragment of an EST-related nucleic acid is hybridized to the single stranded phagemids. Preferably, the fragment comprises at least 10, 12, 15, 17, 18, 20, 23, 25, or 28 consecutive nucleotides of the sequences of SEQ ID NOs. 24-811 and 1600-1622.

Hybrids between the biotinylated oligonucleotide and phagemids are isolated by incubating the hybrids with streptavidin coated paramagnetic beads and retrieving the beads with a magnet (Fry *et al.*, *Biotechniques*, 13: 124-131, 1992). Thereafter, the resulting phagemids are released from the beads and converted into double stranded DNA using a primer specific for the 5' EST or consensus contiguated 5'EST sequence used to design the biotinylated oligonucleotide. Alternatively, protocols such as the Gene Trapper kit (Gibco BRL) may be used. The resulting double stranded DNA is transformed into bacteria. Extended cDNAs or full length cDNAs containing the 5' EST or consensus contiguated 5'EST sequence are identified by colony PCR or colony hybridization.

Using any of the above described methods in section III, a plurality of extended cDNAs containing full-length protein coding sequences or portions of the protein coding sequences may be provided as cDNA libraries for subsequent evaluation of the encoded proteins or use in diagnostic assays as described below.

5

### EXAMPLE 21

#### Full Length cDNAs

The procedures described in Example 19 and 20 were used to obtain extended cDNAs or full length cDNAs derived from 5' ESTs in a variety of tissues. The following list provides a few examples  
10 of cDNAs obtained by these means.

Using this procedure, the full length cDNA of SEQ ID NO:1 (internal identification number 58-34-2-E7-FL2) was obtained. This cDNA encodes the signal peptide MWWFQQGLSFLPSALVIWTSA (SEQ ID NO:2) having a von Heijne score of 5.5.

Using this approach, the full length cDNA of SEQ ID NO:3 (internal identification number 48-  
15 19-3-G1-FL1) was obtained. This cDNA encodes the signal peptide MKKVLLLLITAILAVAVG (SEQ ID NO: 4) having a von Heijne score of 8.2.

The full length cDNA of SEQ ID NO:5 (internal identification number 58-35-2-F10-FL2) was also obtained using this procedure. This cDNA encodes a signal peptide LWLLFFLVTAIHA (SEQ ID NO:6) having a von Heijne score of 10.7.

20 Furthermore, the polypeptides encoded by the extended or full-length cDNAs may be screened for the presence of known structural or functional motifs or for the presence of signatures, small amino acid sequences which are well conserved amongst the members of a protein family. The results obtained for the polypeptides encoded by a few full-length cDNAs derived from 5'ESTs that were screened for the presence of known protein signatures and motifs using the Proscan software from the GCG  
25 package and the Prosite 15.0 database are provided below.

The protein of SEQ ID NO: 8 encoded by the full-length cDNA SEQ ID NO: 7 (internal designation 78-8-3-E6-CL0\_1C) and expressed in adult prostate belong to the phosphatidylethanolamine-binding protein from which it exhibits the characteristic PROSITE signature from positions 90 to 112. Proteins from this widespread family, from nematodes to fly,  
30 yeast, rodent and primate species, bind hydrophobic ligands such as phospholipids and nucleotides. They are mostly expressed in brain and in testis and are thought to play a role in cell growth and/or maturation, in regulation of the sperm maturation, motility and in membrane remodeling. They may act either through signal transduction or through oxidoreduction reactions (for a review see Schoentgen and Jollès, *FEBS Letters*, 369 :22-26 (1995)). Taken together, these data suggest that the  
35 protein of SEQ ID NO: 8 may play a role in cell growth, maturation and in membrane remodeling and/or may be related to male fertility. Thus, these protein may be useful in diagnosing and/or treating cancer, neurodegenerative diseases, and/or disorders related to male fertility and sterility.

The protein of SEQ ID No. 10 encoded by the full-length cDNA SEQ ID NO. 9 (internal designation 108-013-5-O-H9-FLC) shows homologies with a family of lysophospholipases conserved among eukaryotes (yeast, rabbit, rodents and human). In addition, some members of this family exhibit a calcium-independent phospholipase A2 activity (Portilla *et al*, *J. Am. Soc. Nephro.*, 9 :1178-1186 (1998)). All members of this family exhibit the active site consensus GX SXG motif of carboxylesterases that is also found in the protein of SEQ ID NO. 10 (position 54 to 58). In addition, this protein may be a membrane protein with one transmembrane domain as predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, 10 :685-686 (1994)). Taken together, these data suggest that the protein of SEQ ID NO:10 may play a role in fatty acid metabolism, probably as a phospholipase. Thus, this protein or part therein, may be useful in diagnosing and/or treating several disorders including, but not limited to, cancer, diabetes, and neurodegenerative disorders such as Parkinson's and Alzheimer's diseases. It may also be useful in modulating inflammatory responses to infectious agents and/or to suppress graft rejection.

The protein of SEQ ID NO: 12 encoded by the full-length cDNA SEQ ID NO: 11 (internal designation 108-004-5-0-D10-FLC) shows remote homology to a subfamily of beta4-galactosyltransferases widely conserved in animals (human, rodents, cow and chicken). Such enzymes, usually type II membrane proteins located in the endoplasmic reticulum or in the Golgi apparatus, catalyzes the biosynthesis of glycoproteins, glycolipid glycans and lactose. Their characteristic features defined as those of subfamily A in Breton *et al*, *J. Biochem.*, 123:1000-1009 (1998) are pretty well conserved in the protein of SEQ ID NO: 12, especially the region I containing the DVD motif (positions 163-165) thought to be involved either in UDP binding or in the catalytic process itself. In addition, the protein of SEQ ID NO: 12 has the typical structure of a type II protein. Indeed, it contains a short 28-amino-acid-long N-terminal tail, a transmembrane segment from positions 29 to 49 and a large 278-amino-acid-long C-terminal tail as predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, 10 :685-686 (1994)). Taken together, these data suggest that the protein of SEQ ID NO: 12 may play a role in the biosynthesis of polysaccharides, and of the carbohydrate moieties of glycoproteins and glycolipids and/or in cell-cell recognition. Thus, this protein may be useful in diagnosing and/or treating several types of disorders including, but not limited to, cancer, atherosclerosis, cardiovascular disorders, autoimmune disorders and rheumatic diseases including rheumatoid arthritis.

The protein of SEQ ID NO: 14 encoded by the full-length cDNA SEQ ID NO: 13 (internal designation 108-009-5-0-A2-FLC) shows extensive homology to the bZIP family of transcription factors, and especially to the human protein (Lu *et al*, *Mol. Cell. Biol.*, 17 :5117-5126 (1997)). The match include the whole bZIP domain composed of a basic DNA-binding domain and of a leucine zipper allowing protein dimerization. The basic domain is conserved in the protein of SEQ ID NO: 14 as shown by the characteristic PROSITE signature (positions 224-237) except for a conservative substitution of a glutamic acid with an aspartic acid in position 233. The typical

PROSITE signature for leucine zipper is also present (positions 259 to 280). Taken together, these data suggest that the protein of SEQ ID NO: 14 may bind to DNA, hence regulating gene expression as a transcription factor. Thus, this protein may be useful in diagnosing and/or treating several types of disorders including, but not limited to, cancer.

5 Bacterial clones containing plasmids containing the full length cDNAs described above are presently stored in the inventor's laboratories under the internal identification numbers provided above. The inserts may be recovered from the deposited materials by growing an aliquot of the appropriate bacterial clone in the appropriate medium. The plasmid DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale  
10 alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the insertion. The PCR product which corresponds to the cDNA insert  
15 can then be manipulated using standard cloning techniques familiar to those skilled in the art.

#### **V. Expression of Proteins or Polypeptides Encoded by EST-related nucleic acids or Fragments thereof**

EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-  
20 related nucleic acids, and fragments of positional segments of EST-related nucleic acids may be used to express the polypeptides which they encode. In particular, they may be used to express EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides, or fragments of positional segments of EST-related polypeptides. In some embodiments, the EST-related nucleic acids, positional segments of EST-related nucleic acids, and fragments of positional  
25 segments of EST-related nucleic acids may be used to express the full polypeptide (*i.e.* the signal peptide and the mature polypeptide) of a secreted protein, the mature protein (*i.e.* the polypeptide generated after cleavage of the signal peptide), or the signal peptide of a secreted protein. If desired, nucleic acids encoding the signal peptide may be used to facilitate secretion of the expressed protein. It will be appreciated that a plurality of EST-related nucleic acids, fragments of EST-related nucleic acids,  
30 positional segments of EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids may be simultaneously cloned into expression vectors to create an expression library for analysis of the encoded proteins as described below.

#### **EXAMPLE 22**

35 Expression of the Proteins Encoded by the Genes Corresponding to the  
5'ESTs or Consensus Contigated 5' ESTs



To express their encoded proteins, the EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids are cloned into a suitable expression vector. In some instances, nucleic acids encoding EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides or fragments of positional segments of EST-related polypeptides may be cloned into a suitable expression vector.

In some embodiments, the nucleic acids inserted into the expression vector may comprise the coding sequence of a sequence selected from the group consisting of SEQ ID NOs. 24-811. In other embodiments, the nucleic acids inserted into the expression vector may comprise the full coding sequence (*i.e.* the nucleotides encoding the signal peptide and the mature polypeptide) of one of SEQ ID Nos. 766-792. In some embodiments, the nucleic acid inserted into the expression vector may comprise the nucleotides of one of the sequences of SEQ ID Nos. 766-792 which encode the mature polypeptide (*i.e.* the nucleotides encoding the polypeptide generated after cleavage of the signal peptide). In further embodiments, the nucleic acids inserted into the expression vector may comprise the nucleotides of 24-728 and 766-792 which encode the signal peptide to facilitate secretion of the expressed protein. The nucleic acids inserted into the expression vectors may also contain sequences upstream of the sequences encoding the signal peptide, such as sequences which regulate expression levels or sequences which confer tissue specific expression.

The nucleic acid inserted into the expression vector may encode a polypeptide comprising the one of the sequences of SEQ ID Nos. 812-1599. In some embodiments, the nucleic acid inserted into the expression vector may encode the full polypeptide sequence (*i.e.* the signal peptide and the mature polypeptide) included in one of SEQ ID Nos. 1554-1580. In other embodiments, the nucleic acid inserted into the expression vector may encode the mature polypeptide (*i.e.* the polypeptide generated after cleavage of the signal peptide) included in one of the sequences of SEQ ID Nos. 1554-1580. In further embodiments, the nucleic acids inserted into the expression vector may encode the signal peptide included in one of the sequences of 812-1516 and 1554-1580.

The nucleic acid encoding the protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The expression vector may be any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, *et al.*, U.S. Patent No. 5,082,767.

The following is provided as one exemplary method to express the proteins encoded by the nucleic acids described above. In some instances the nucleic acid encoding the protein or polypeptide to

- be expressed includes a methionine initiation codon and a polyA signal. If the nucleic acid encoding the polypeptide to be expressed lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the nucleic acid encoding the protein or polypeptide to be expressed lacks a polyA signal, this sequence can
- 5 be added to the construct by, for example, splicing out the polyA signal from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the *gag* gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene.
- 10 The nucleic acid encoding the polypeptide to be expressed is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the nucleic acid encoding the protein or polypeptide to be expressed and containing restriction endonuclease sequences for Pst I incorporated into the 5' primer and BglII at the 5' end of 3' primer, taking care to ensure that the nucleic acid encoding the protein or polypeptide to be expressed is correctly positioned with respect to the poly A signal. The purified
- 15 fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A signal and digested with BglII.

- The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification.
- 20 Positive transfectants are selected after growing the transfected cells in 600 µg/ml G418 (Sigma, St. Louis, Missouri).

- Alternatively, the nucleic acid encoding the protein or polypeptide to be expressed may be cloned into pED6dpc2. The resulting pED6dpc2 constructs may be transfected into a suitable host cell, such as COS 1 cells. Methotrexate resistant cells are selected and expanded. The expressed protein or
- 25 polypeptide may be isolated, purified, or enriched as described above.

- To confirm expression of the desired protein or polypeptide, the proteins or polypeptides produced by cells containing a vector with a nucleic acid insert encoding the protein or polypeptide are compared to those lacking such an insert. The expressed proteins are detected using techniques familiar to those skilled in the art such as Coomassie blue or silver staining or using antibodies against the protein
- 30 or polypeptide encoded by the nucleic acid insert. Antibodies capable of specifically recognizing the protein of interest may be generated using synthetic 15-mer peptides having a sequence encoded by the appropriate nucleic acid. The synthetic peptides are injected into mice to generate antibody to the polypeptide encoded by the nucleic acid.

- If the proteins or polypeptides encoded by the nucleic acid inserts are secreted, medium
- 35 prepared from the host cells or organisms containing an expression vector which contains a nucleic acid insert encoding the desired protein or polypeptide is compared to medium prepared from the control cells or organism. The presence of a band in medium from the cells containing the nucleic acid insert which

is absent from preparations from the control cells indicates that the protein or polypeptide encoded by the nucleic acid insert is being expressed and secreted. Generally, the band corresponding to the protein encoded by the nucleic acid insert will have a mobility near that expected based on the number of amino acids in the open reading frame of the nucleic acid insert. However, the band may have a mobility  
5 different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Alternatively, if the protein expressed from the above expression vectors does not contain sequences directing its secretion, the proteins expressed from host cells containing an expression vector with an insert encoding a secreted protein or portion thereof can be compared to the proteins expressed  
10 in control host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the desired protein or portion thereof is being expressed. Generally, the band will have the mobility expected for the secreted protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications  
15 such as glycosylation, ubiquitination, or enzymatic cleavage.

The expressed protein or polypeptide may be purified, isolated or enriched using a variety of methods. In some methods, the protein or polypeptide may be secreted into the culture medium via a native signal peptide or a heterologous signal peptide operably linked thereto. In some methods, the protein or polypeptide may be linked to a heterologous polypeptide which facilitates its isolation,  
20 purification, or enrichment such as a nickel binding polypeptide. The protein or polypeptide may also be obtained by gel electrophoresis, ion exchange chromatography, size chromatography, hplc, salt precipitation, immunoprecipitation, a combination of any of the preceding methods, or any of the isolation, purification, or enrichment techniques familiar to those skilled in the art.

The protein encoded by the nucleic acid insert may also be purified using standard  
25 immunochromatography techniques using immunoaffinity chromatography with antibodies directed against the encoded protein or polypeptide as described in more detail below. If antibody production is not possible, the nucleic acid insert encoding the desired protein or polypeptide may be incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies, the coding sequence of the nucleic acid insert is ligated in frame with the gene encoding the  
30 other half of the chimera. The other half of the chimera may be  $\beta$ -globin or a nickel binding polypeptide. A chromatography matrix having antibody to  $\beta$ -globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites may be engineered between the  $\beta$ -globin gene or the nickel binding polypeptide and the extended cDNA or portion thereof. Thus, the two polypeptides of the chimera may be separated from one another by protease digestion.

35 One useful expression vector for generating  $\beta$ -globin chimerics is pSG5 (Stratagene), which encodes rabbit  $\beta$ -globin. Intron II of the rabbit  $\beta$ -globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of

expression. These techniques as described are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, (*Basic Methods in Molecular Biology*, L.G. Davis, M.D. Digner, and J.F. Battey, ed., Elsevier Press, NY, 1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using *in vitro* translation systems such as the *In vitro* Express™ Translation Kit (Stratagene).

Following expression and purification of the proteins or polypeptides encoded by the nucleic acid inserts, the purified proteins may be tested for the ability to bind to the surface of various cell types as described in Example 23 below. It will be appreciated that a plurality of proteins expressed from these nucleic acid inserts may be included in a panel of proteins to be simultaneously evaluated for the activities specifically described below, as well as other biological roles for which assays for determining activity are available.

### EXAMPLE 23

#### Analysis of Secreted Proteins to Determine Whether they Bind to the Cell Surface

The EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids, fragments of positional segments of EST-related nucleic acids, nucleic acids encoding the EST-related polypeptides, nucleic acids encoding fragments of the EST-related polypeptides, nucleic acids encoding positional segments of EST-related polypeptides, or nucleic acids encoding fragments of positional segments of EST-related polypeptides are cloned into expression vectors such as those described in Example 22. The encoded proteins or polypeptides are purified, isolated, or enriched as described above. Following purification, isolation, or enrichment, the proteins or polypeptides are labeled using techniques known to those skilled in the art. The labeled proteins or polypeptides are incubated with cells or cell lines derived from a variety of organs or tissues to allow the proteins to bind to any receptor present on the cell surface. Following the incubation, the cells are washed to remove non-specifically bound proteins or polypeptides. The specifically bound labeled proteins or polypeptides are detected by autoradiography. Alternatively, unlabeled proteins or polypeptides may be incubated with the cells and detected with antibodies having a detectable label, such as a fluorescent molecule, attached thereto.

Specificity of cell surface binding may be analyzed by conducting a competition analysis in which various amounts of unlabeled protein or polypeptide are incubated along with the labeled protein or polypeptide. The amount of labeled protein or polypeptide bound to the cell surface decreases as the amount of competitive unlabeled protein or polypeptide increases. As a control, various amounts of an unlabeled protein or polypeptide unrelated to the labeled protein or polypeptide is included in some binding reactions. The amount of labeled protein or polypeptide bound to the cell surface does not decrease in binding reactions containing increasing amounts of unrelated unlabeled protein, indicating that the protein or polypeptide encoded by the nucleic acid binds specifically to the cell surface.

As discussed above, human proteins have been shown to have a number of important physiological effects and, consequently, represent a valuable therapeutic resource. The human proteins or polypeptides made as described above may be evaluated to determine their physiological activities as described below.

5

#### EXAMPLE 24

##### Assaying the Expressed Proteins or Polypeptides for Cytokine, Cell Proliferation or Cell Differentiation Activity

As discussed above, some human proteins act as cytokines or may affect cellular proliferation or  
10 differentiation. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein or polypeptide of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M<sup>+</sup> (preB  
15 M<sup>+</sup>), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7c and CMK. The proteins or polypeptides prepared as described above may be evaluated for their ability to regulate T cell or thymocyte proliferation in assays such as those described above or in the following references: *Current Protocols in Immunology*, Ed. by J.E. Coligan *et al.*, Greene Publishing Associates and Wiley-Interscience; Takai *et al. J. Immunol.* 137:3494-3500, 1986., Bertagnolli *et al. J. Immunol.* 145:1706-1712, 1990.,  
20 Bertagnolli *et al., Cellular Immunology* 133:327-341, 1991. Bertagnolli, *et al. J. Immunol.* 149:3778-3783, 1992; Bowman *et al., J. Immunol.* 152:1756-1761, 1994.

In addition, numerous assays for cytokine production and/or the proliferation of spleen cells, lymph node cells and thymocytes are known. These include the techniques disclosed in *Current Protocols in Immunology*. J.E. Coligan *et al.* Eds., 1:3.12.1-3.12.14, John Wiley and Sons, Toronto.  
25 1994; and Schreiber, R.D. In *Current Protocols in Immunology.*, *supra* 1 : 6.8.1-6.8.8.

The proteins or polypeptides prepared as described above may also be assayed for the ability to regulate the proliferation and differentiation of hematopoietic or lymphopoietic cells. Many assays for such activity are familiar to those skilled in the art, including the assays in the following references: Bottomly *et al.*, In *Current Protocols in Immunology.*, *supra* 1 : 6.3.1-6.3.12.,; deVries *et al., J. Exp.*  
30 *Med.* 173:1205-1211, 1991; Moreau *et al., Nature* 36:690-692, 1988; Greenberger *et al., Proc. Natl. Acad. Sci. U.S.A.* 80:2931-2938, 1983; Nordan, R., In *Current Protocols in Immunology.*, *supra* 1 : 6.6.1-6.6.5; Smith *et al., Proc. Natl. Acad. Sci. U.S.A.* 83:1857-1861, 1986; Bennett *et al* in *Current Protocols in Immunology supra* 1 : 6.15.1; Ciarletta *et al* In *Current Protocols in Immunology. supra* 1 : 6.13.1.

35 The proteins or polypeptides prepared as described above may also be assayed for their ability to regulate T-cell responses to antigens. Many assays for such activity are familiar to those skilled in the art, including the assays described in the following references: Chapter 3 (*In vitro* Assays for Mouse

Lymphocyte Function), Chapter 6 (Cytokines and Their Cellular Receptors) and Chapter 7, (Immunologic Studies in Humans) in *Current Protocols in Immunology supra*; Weinberger *et al.*, *Proc. Natl. Acad. Sci. USA* 77:6091-6095, 1980; Weinberger *et al.*, *Eur. J. Immunol.* 11:405-411, 1981; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988.

5        Those proteins or polypeptides which exhibit cytokine, cell proliferation, or cell differentiation activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which induction of cell proliferation or differentiation is beneficial. Alternatively, as described in more detail below, nucleic acids encoding these proteins or polypeptides or nucleic acids regulating the expression of these proteins or polypeptides may be introduced into appropriate host cells to increase or decrease the  
10    expression of the proteins or polypeptides as desired.

### EXAMPLE 25

#### Assaying the Expressed Proteins or Polypeptides for Activity as Immune System Regulators

15        The proteins or polypeptides prepared as described above may also be evaluated for their effects as immune regulators. For example, the proteins or polypeptides may be evaluated for their activity to influence thymocyte or splenocyte cytotoxicity. Numerous assays for such activity are familiar to those skilled in the art including the assays described in the following references: Chapter 3 (*In vitro* Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic studies in Humans) in *Current  
20 Protocols in Immunology*, J.E. Coligan *et al.* Eds, Greene Publishing Associates and Wiley-Interscience; Herrmann *et al.*, *Proc. Natl. Acad. Sci. USA* 78:2488-2492, 1981; Herrmann *et al.*, *J. Immunol.* 128:1968-1974, 1982; Handa *et al.*, *J. Immunol.* 135:1564-1572, 1985; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988; Bowman *et al.*, *J. Virology* 61:1992-1998; Bertagnolli *et al. Cell. Immunol.* 133:327-341, 1991; Brown *et al.*, *J. Immunol.* 153:3079-3092,  
25    1994.

      The proteins or polypeptides prepared as described above may also be evaluated for their effects on T-cell dependent immunoglobulin responses and isotype switching. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Maliszewski, *J. Immunol.* 144:3028-3033, 1990; Mond *et al.* in *Current Protocols in Immunology*, 1 :  
30    3.8.1-3.8.16, *supra*.

      The proteins or polypeptides prepared as described above may also be evaluated for their effect on immune effector cells, including their effect on Th1 cells and cytotoxic lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Chapter 3 (*In vitro* Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter  
35    7 (Immunologic Studies in Humans) in *Current Protocols in Immunology, supra*; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988; Bertagnolli *et al.*, *J. Immunol.* 149:3778-3783, 1992.

The proteins or polypeptides prepared as described above may also be evaluated for their effect on dendritic cell mediated activation of naive T-cells. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Guery *et al.*, *J. Immunol.* 134:536-544, 1995; Inaba *et al.*, *J. Exp. Med.* 173:549-559, 1991; Macatonia *et al.*, *J. Immunol.* 154:5071-5079, 1995; Porgador *et al.* *J. Exp. Med.* 182:255-260, 1995; Nair *et al.*, *J. Virol.* 67:4062-4069, 1993; Huang *et al.*, *Science* 264:961-965, 1994; Macatonia *et al.* *J. Exp. Med.* 169:1255-1264, 1989; Bhardwaj *et al.*, *Journal of Clinical Investigation* 94:797-807, 1994; and Inaba *et al.*, *J. Exp. Med.* 172:631-640, 1990.

The proteins or polypeptides prepared as described above may also be evaluated for their influence on the lifetime of lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Darzynkiewicz *et al.*, *Cytometry* 13:795-808, 1992; Gorczyca *et al.*, *Leukemia* 7:659-670, 1993; Gorczyca *et al.*, *Cancer Res.* 53:1945-1951, 1993; Itoh *et al.*, *Cell* 66:233-243, 1991; Zacharchuk, *J. Immunol.* 145:4037-4045, 1990; Zamai *et al.*, *Cytometry* 14:891-897, 1993; Gorczyca *et al.*, *Int. J. Oncol.* 1:639-648, 1992.

The proteins or polypeptides prepared as described above may also be evaluated for their influence on early steps of T-cell commitment and development. Numerous assays for such activity are familiar to those skilled in the art, including without limitation the assays disclosed in the following references: Antica *et al.*, *Blood* 84:111-117, 1994; Fine *et al.*, *Cell. Immunol.* 155:111-122, 1994; Galy *et al.*, *Blood* 85:2770-2778, 1995; Toki *et al.*, *Proc. Nat. Acad. Sci. USA* 88:7548-7551, 1991.

Those proteins or polypeptides which exhibit activity as immune system regulators activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of immune activity is beneficial. For example, the protein or polypeptide may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using the protein or polypeptide including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, *Leishmania* spp., *Plasmodium* and various fungal infections such as candidiasis. Of course, in this regard, a protein or polypeptide may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Alternatively, the proteins or polypeptides prepared as described above may be used in treatment of autoimmune disorders including, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein or polypeptide may also be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic

asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using the protein or polypeptide.

Using the proteins or polypeptides of the invention it may also be possible to regulate immune responses either up or down. Down regulation may involve inhibiting or blocking an immune response  
5 already in progress or may involve preventing the induction of an immune response. The functions of activated T-cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active non-antigen-specific process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression  
10 in that it is generally antigen-specific and persists after the end of exposure to the tolerizing agent. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions, such as, for example, B7 costimulation), e.g., preventing high level  
15 lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks  
20 interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation, can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen  
25 function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the  
30 function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig  
35 fusion proteins *in vivo* as described in Lenschow *et al.*, *Science* 257:789-792 (1992) and Turka *et al.*, *Proc. Natl. Acad. Sci USA*, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul *et al.*,



*Fundamental Immunology*, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor/ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which potentially involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/pr/pr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in OD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may involve either enhancing an existing immune response or eliciting an initial immune response as shown by the following examples. For instance, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory form of B lymphocyte antigens systemically.

Alternatively, antiviral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells *in vitro* with viral antigen-pulsed APCs either expressing the proteins or polypeptides described above or together with a stimulatory form of the protein or polypeptide and reintroducing the *in vitro* primed T cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to T cells *in vivo*, thereby activating the T cells.

In another application, upregulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with one of the above-described nucleic acids encoding a protein or polypeptide can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected *ex vivo* with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor

cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the protein or polypeptide encoded by the nucleic acids described above having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules can be transfected with nucleic acids encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I  $\alpha$  chain and  $\beta_2$  microglobulin or an MHC class II  $\alpha$  chain and an MHC class II  $\beta$  chain to thereby express MHC class I or MHC class II proteins on the cell surface, respectively. Expression of the appropriate MHC class I or class II molecules in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a nucleic acid encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a protein or polypeptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject. Alternatively, as described in more detail below, nucleic acids encoding these immune system regulator proteins or polypeptides or nucleic acids regulating the expression of such proteins or polypeptides may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

## EXAMPLE 26

### Assaying the Expressed Proteins or Polypeptides for Hematopoiesis Regulating Activity

The proteins or polypeptides encoded by the nucleic acids described above may also be evaluated for their hematopoiesis regulating activity. For example, the effect of the proteins or polypeptides on embryonic stem cell differentiation may be evaluated. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Johansson *et al. Cell. Biol.* 15:141-151, 1995; Keller *et al., Mol. Cell. Biol.* 13:473-486, 1993; McClanahan *et al., Blood* 81:2903-2915, 1993.

The proteins or polypeptides encoded by the nucleic acids described above may also be evaluated for their influence on the lifetime of stem cells and stem cell differentiation. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Freshney, M.G. Methylcellulose Colony Forming Assays, in Culture of Hematopoietic Cells. R.I. Freshney, *et al.* Eds. pp. 265-268, Wiley-Liss, Inc., New York, NY. 1994; Hirayama *et al., Proc. Natl. Acad. Sci. USA* 89:5907-5911, 1992; McNiece, I.K. and Briddell, R.A. Primitive Hematopoietic Colony Forming Cells with High Proliferative Potential, in Culture of Hematopoietic Cells. supra;

Neben *et al.*, *Experimental Hematology* 22:353-359, 1994; Ploemacher, R.E. Cobblestone Area Forming Cell Assay, In Culture of Hematopoietic Cells. *supra*; Spooncer, E., Dexter, M. and Allen, T. Long Term Bone Marrow Cultures in the Presence of Stromal Cells, in Culture of Hematopoietic Cells *supra*; and Sutherland, H.J. Long Term Culture Initiating Cell Assay, in Culture of Hematopoietic Cells. *supra*.

5        Those proteins or polypeptides which exhibit hematopoiesis regulatory activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of hematopoiesis is beneficial. For example, a protein or polypeptide of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates  
10 involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (*i.e.*, traditional CSF activity) useful, for  
15 example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-  
20 mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (*i.e.*, in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as  
25 normal cells or genetically manipulated for gene therapy. Alternatively, as described in more detail below, nucleic acids encoding these proteins or polypeptides or nucleic acids regulating the expression of these proteins or polypeptides may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

30

#### EXAMPLE 27

##### Assaying the Expressed Proteins or Polypeptides for Regulation of Tissue Growth

The proteins or polypeptides encoded by the nucleic acids described above may also be evaluated for their effect on tissue growth. Numerous assays for such activity are familiar to those  
35 skilled in the art, including the assays disclosed in International Patent Publication No. WO95/16035, International Patent Publication No. WO95/05846 and International Patent Publication No. WO91/07491.

Assays for wound healing activity include, without limitation, those described in: Winter, *Epidermal Wound Healing*, pps. 71-112 (Maibach, H1 and Rovee, DT, eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

Those proteins or polypeptides which are involved in the regulation of tissue growth may then  
5 be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of tissue growth is beneficial. For example, a protein or polypeptide may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein or polypeptide encoded by the nucleic acids described above which induces cartilage  
10 and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein or polypeptide of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. *De novo* bone synthesis induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection  
15 induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein or polypeptide of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or  
20 osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the proteins or polypeptides encoded by the nucleic acids described above is tendon/ligament formation. A protein or  
25 polypeptide encoded by the nucleic acids described above, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved  
30 fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. *De novo* tendon/ligament-like tissue formation induced by a protein or polypeptide of the present invention contributes to the repair of tendon or ligaments defects of congenital, traumatic or other origin and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The proteins or polypeptides of the present invention may provide an environment to attract tendon- or  
35 ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The proteins or polypeptides of the

invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The therapeutic compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The proteins or polypeptides of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, *i.e.*, for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein or polypeptide may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein or polypeptide of the invention.

Proteins or polypeptides of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein or polypeptide of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium) muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to generate. A protein or polypeptide of the invention may also exhibit angiogenic activity.

A protein or polypeptide of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein or polypeptide of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Alternatively, as described in more detail below, nucleic acids encoding tissue growth regulating activity proteins or polypeptides or nucleic acids regulating the expression of such proteins or polypeptides may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

## EXAMPLE 28

### Assaying the Expressed Proteins or Polypeptides for Regulation of Reproductive Hormones

The proteins or polypeptides of the present invention may also be evaluated for their ability to regulate reproductive hormones, such as follicle stimulating hormone. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Vale *et al.*, *Endocrinol.* 91:562-572, 1972; Ling *et al.*, *Nature* 321:779-782, 1986; Vale *et al.*, *Nature* 321:776-779, 1986; Mason *et al.*, *Nature* 318:659-663, 1985; Forage *et al.*, *Proc. Natl. Acad. Sci. USA* 83:3091-3095, 1986. Chapter 6.12 in *Current Protocols in Immunology*, J.E. Coligan *et al.* Eds. Greene Publishing Associates and Wiley-Interscience; Taub *et al.* *J. Clin. Invest.* 95:1370-1376, 1995; Lind *et al.* *APMIS* 103:140-146, 1995; Muller *et al.* *Eur. J. Immunol.* 25:1744-1748; Gruber *et al.* *J. Immunol.* 152:5860-5867, 1994; Johnston *et al.*, *J. Immunol.* 153:1762-1768, 1994.

Those proteins or polypeptides which exhibit activity as reproductive hormones or regulators of cell movement may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of reproductive hormones are beneficial. For example, a protein or polypeptide may exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of FSH. Thus, a protein or polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein or polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-B group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, United States Patent 4,798,885. A protein or polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

Alternatively, as described in more detail below, nucleic acids encoding reproductive hormone regulating activity proteins or polypeptides or nucleic acids regulating the expression of such proteins or polypeptides may be introduced into appropriate host cells to increase or decrease the expression of the proteins or polypeptides as desired.

## EXAMPLE 29

### Assaying the Expressed Proteins or Polypeptides For Chemotactic/Chemokinetic Activity

The proteins or polypeptides of the present invention may also be evaluated for chemotactic/chemokinetic activity. For example, a protein or polypeptide of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins or polypeptides can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins or

polypeptides provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

5 A protein or polypeptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or polypeptide has the ability to directly stimulate directed movement of cells. Whether a particular protein or polypeptide has chemotactic activity for a population of cells can be readily determined by employing such protein or polypeptide in any known assay for cell chemotaxis.

10 The activity of a protein or polypeptide of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins or polypeptides that induce or prevent chemotaxis) consist of assays that measure the ability of a protein or polypeptide to induce the migration of cells across a membrane as well as the ability of a protein or polypeptide to induce the  
15 adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: *Current Protocols in Immunology*, Ed by J.E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience, Chapter 6.12: 6.12.1-6.12.28; Taub *et al. J. Clin. Invest.* 95:1370-1376, 1995; Lind *et al. APMIS* 103:140-146, 1995; Mueller *et al., Eur. J. Immunol.* 25:1744-1748; Gruber *et al. J. Immunol.*  
20 152:5860-5867, 1994; Johnston *et al. J. Immunol.*, 153:1762-1768, 1994.

### EXAMPLE 30

#### Assaying the Expressed Proteins or Polypeptides for Regulation of Blood Clotting

The proteins or polypeptides of the present invention may also be evaluated for their effects on  
25 blood clotting. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Linet *et al., J. Clin. Pharmacol.* 26:131-140, 1986; Burdick *et al., Thrombosis Res.* 45:413-419, 1987; Humphrey *et al., Fibrinolysis* 5:71-79 (1991); Schaub, *Prostaglandins* 35:467-474, 1988.

Those proteins or polypeptides which are involved in the regulation of blood clotting may then  
30 be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of blood clotting is beneficial. For example, a protein or polypeptide of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein or polypeptide is expected to be useful in treatment of various coagulations disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other  
35 causes. A protein or polypeptide of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as infarction of cardiac and central nervous system vessels (e.g., stroke)). Alternatively, as described in

more detail below, nucleic acids encoding blood clotting activity proteins or polypeptides or nucleic acids regulating the expression of such proteins or polypeptides may be introduced into appropriate host cells to increase or decrease the expression of the proteins or polypeptides as desired.

5

**EXAMPLE 31**

Assaying the Expressed Proteins or Polypeptides for Involvement in  
Receptor/Ligand Interactions

The proteins or polypeptides of the present invention may also be evaluated for their involvement in receptor/ligand interactions. Numerous assays for such involvement are familiar to those skilled in the art, including the assays disclosed in the following references: Chapter 7. 7.28.1-7.28.22) in *Current Protocols in Immunology*, J.E. Coligan *et al.* Eds. Greene Publishing Associates and Wiley-Interscience; Takai *et al.*, *Proc. Natl. Acad. Sci. USA* 84:6864-6868, 1987; Bierer *et al.*, *J. Exp. Med.* 168:1145-1156, 1988; Rosenstein *et al.*, *J. Exp. Med.* 169:149-160, 1989; Stoltenborg *et al.*, *J. Immunol. Methods* 175:59-68, 1994; Stitt *et al.*, *Cell* 80:661-670, 1995; Gyuris *et al.*, *Cell* 75:791-803, 1993.

For example, the proteins or polypeptides of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein or polypeptide of the present invention (including, without limitation, fragments of receptors and ligands) may be useful as inhibitors of receptor/ligand interactions. Alternatively, as described in more detail below, nucleic acids encoding proteins or polypeptides involved in receptor/ligand interactions or nucleic acids regulating the expression of such proteins or polypeptides may be introduced into appropriate host cells to increase or decrease the expression of the proteins or polypeptides as desired.

**EXAMPLE 32**

30

Assaying the Proteins or Polypeptides for Anti-Inflammatory Activity

The proteins or polypeptides of the present invention may also be evaluated for anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins or polypeptides exhibiting such activities can be used to treat inflammatory conditions



including chronic or acute conditions, including without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine- or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or  
5 resulting from over production of cytokines such as TNF or IL-1. Proteins or polypeptides of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Alternatively, as described in more detail below, nucleic acids encoding anti-inflammatory activity proteins or polypeptides or nucleic acids regulating the expression of such proteins or polypeptides may be introduced into appropriate host cells to increase or decrease the expression of the  
10 proteins or polypeptides as desired.

### EXAMPLE 33

#### Assaying the Expressed Proteins or Polypeptides for Tumor Inhibition Activity

The proteins or polypeptides of the present invention may also be evaluated for tumor inhibition  
15 activity. In addition to the activities described above for immunological treatment or prevention of tumors, a protein or polypeptide of the invention may exhibit other anti-tumor activities. A protein or polypeptide may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein or polypeptide may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for  
20 example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth. . Alternatively, as described in more detail below, nucleic acids encoding proteins or polypeptides with tumor inhibition activity or nucleic acids regulating the expression of such proteins or polypeptides may be introduced into appropriate host cells to increase or decrease the  
25 expression of the proteins or polypeptides as desired.

A protein or polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color,  
30 skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting  
35 behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem

cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein. Alternatively, as described in more detail below, nucleic acids encoding proteins or polypeptides involved in any of the above mentioned activities or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins or polypeptides as desired.

10

### EXAMPLE 34

#### Identification of Proteins or Polypeptides which Interact with Proteins or Polypeptides of the Present Invention

Proteins or polypeptides which interact with the proteins or polypeptides of the present invention, such as receptor proteins, may be identified using two hybrid systems such as the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the manual accompanying the kit, nucleic acids encoding the proteins or polypeptides of the present invention, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. cDNAs in a cDNA library which encode proteins or polypeptides which might interact with the proteins or polypeptides of the present invention are inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain plasmids encoding proteins or polypeptides which interact with the proteins or polypeptides of the present invention.

Alternatively, the system described in Lustig *et al.*, *Methods in Enzymology* 283: 83-99 (1997) may be used for identifying molecules which interact with the proteins or polypeptides of the present invention. In such systems, *in vitro* transcription reactions are performed on a pool of vectors containing nucleic acid inserts which encode the proteins or polypeptides of the present invention. The nucleic acid inserts are cloned downstream of a promoter which drives *in vitro* transcription. The resulting pools of mRNAs are introduced into *Xenopus laevis* oocytes. The oocytes are then assayed for a desired activity.

Alternatively, the pooled *in vitro* transcription products produced as described above may be translated *in vitro*. The pooled *in vitro* translation products can be assayed for a desired activity or for interaction with a known protein or polypeptide.

Proteins, polypeptides or other molecules interacting with proteins or polypeptides of the present invention can be found by a variety of additional techniques. In one method, affinity columns containing the protein or polypeptide of the present invention can be constructed. In some versions, of this method the affinity column contains chimeric proteins in which the protein or polypeptide of the present invention is fused to glutathione S-transferase. A mixture of cellular proteins or pool of expressed proteins as described above and is applied to the affinity column. Molecules interacting with the protein or polypeptide attached to the column can then be isolated and analyzed on 2-D electrophoresis gel as described in Ramunsen *et al. Electrophoresis*, 18, 588-598 (1997). Alternatively, the molecules retained on the affinity column can be purified by electrophoresis based methods and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

Molecules interacting with the proteins or polypeptides of the present invention can also be screened by using an Optical Biosensor as described in Edwards & Leatherbarrow, *Analytical Biochemistry*, 246, 1-6 (1997). The main advantage of the method is that it allows the determination of the association rate between the protein or polypeptide and other interacting molecules. Thus, it is possible to specifically select interacting molecules with a high or low association rate. Typically a target molecule is linked to the sensor surface (through a carboxymethyl dextran matrix) and a sample of test molecules is placed in contact with the target molecules. The binding of a test molecule to the target molecule causes a change in the refractive index and/ or thickness. This change is detected by the Biosensor provided it occurs in the evanescent field (which extends a few hundred nanometers from the sensor surface). In these screening assays, the target molecule can be one of the proteins or polypeptides of the present invention and the test sample can be a collection of proteins, polypeptides or other molecules extracted from tissues or cells, a pool of expressed proteins, combinatorial peptide and/ or chemical libraries, or phage displayed peptides. The tissues or cells from which the test molecules are extracted can originate from any species.

In other methods, a target protein or polypeptide is immobilized and the test population is a collection of unique proteins or polypeptides of the present invention.

To study the interaction of the proteins or polypeptides of the present invention with drugs, the microdialysis coupled to HPLC method described by Wang *et al.*, *Chromatographia*, 44, 205-208(1997) or the affinity capillary electrophoresis method described by Busch *et al.*, *J. Chromatogr.* 777:311-328 (1997) can be used.

The system described in U.S. Patent No. 5,654,150 may also be used to identify molecules which interact with the proteins or polypeptides of the present invention. In this system, pools of nucleic acids encoding the proteins or polypeptides of the present invention are transcribed and translated *in vitro* and the reaction products are assayed for interaction with a known polypeptide or antibody.

It will be appreciated by those skilled in the art that the proteins or polypeptides of the present invention may be assayed for numerous activities in addition to those specifically enumerated above.

For example, the expressed proteins or polypeptides may be evaluated for applications involving control and regulation of inflammation, tumor proliferation or metastasis, infection, or other clinical conditions. In addition, the proteins or polypeptides may be useful as nutritional agents or cosmetic agents.

The proteins or polypeptides of the present invention may be used to generate antibodies  
5 capable of specifically binding to the proteins or polypeptides of the present invention. The antibodies may be monoclonal antibodies or polyclonal antibodies. As used herein, "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where a binding domain is formed from the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution  
10 complementary to the features of an antigenic determinant of an antigen., which allows an immunological reaction with the antigen. Antibodies include recombinant proteins comprising the binding domains, as wells as fragments, including Fab, Fab', F(ab)<sub>2</sub>, and F(ab')<sub>2</sub> fragments.

As used herein, an "antigenic determinant" is the portion of an antigen molecule, that determines the specificity of the antigen-antibody reaction. An "epitope" refers to an antigenic  
15 determinant of a polypeptide. An epitope can comprise as few as 3 amino acids in a spatial conformation which is unique to the epitope. Generally an epitope consists of at least 6 such amino acids, and more usually at least 8-10 such amino acids. Methods for determining the amino acids which make up an epitope include x-ray crystallography, 2-dimensional nuclear magnetic resonance, and epitope mapping e.g. the Pepscan method described by H. Mario Geysen *et al.* 1984. Proc. Natl.  
20 Acad. Sci. U.S.A. 81:3998-4002; PCT Publication No. WO 84/03564; and PCT Publication No. WO 84/03506.

In some embodiments, the antibodies may be capable of specifically binding to a protein or polypeptide encoded by EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids.  
25 In some embodiments, the antibody may be capable of binding an antigenic determinant or an epitope in a protein or polypeptide encoded by EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids.

In other embodiments, the antibodies may be capable of specifically binding to an EST-related  
30 polypeptide, fragment of an EST-related polypeptide, positional segment of an EST-related polypeptide or fragment of a positional segment of an EST-related polypeptide. In some embodiments, the antibody may be capable of binding an antigenic determinant or an epitope in an EST-related polypeptide, fragment of an EST-related polypeptide, positional segment of an EST-related polypeptide or fragment of a positional segment of an EST-related polypeptide.

35 In the case of secreted proteins, the antibodies may be capable of binding a full-length protein encoded by a nucleic acid of the present invention, a mature protein (*i.e.* the protein generated by

cleavage of the signal peptide) encoded by a nucleic acid of the present invention, or a signal peptide encoded by a nucleic acid of the present invention.

### EXAMPLE 35

#### 5                    Production of an Antibody to a Human Polypeptide or Protein

The above described EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or nucleic acids encoding EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides or fragments of positional segments of  
10 EST-related polypeptides are operably linked to promoters and introduced into cells as described above.

In the case of secreted proteins, nucleic acids encoding the full protein (*i.e.* the mature protein and the signal peptide), nucleic acids encoding the mature protein (*i.e.* the protein generated by cleavage of the signal peptide), or nucleic acids encoding the signal peptide are operably linked to promoters and introduced into cells as described above.

15            The encoded proteins or polypeptides are then substantially purified or isolated as described above. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few  $\mu\text{g/ml}$ . Monoclonal or polyclonal antibody to the protein or polypeptide can then be prepared as follows:

#### 1. Monoclonal Antibody Production by Hybridoma Fusion

20            Monoclonal antibody to epitopes of any of the proteins or polypeptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, and Milstein, *Nature* 256:495 (1975) or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein or peptides derived therefrom over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The  
25 spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as Elisa, as  
30 originally described by Engvall, *Meth. Enzymol.* 70:419 (1980). Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.* in *Basic Methods in Molecular Biology* Elsevier, New York. Section 21-2.

#### 2. Polyclonal Antibody Production by Immunization

35            Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein or polypeptide can be prepared by immunizing suitable animals with the expressed protein or peptides derived therefrom, which can be unmodified or modified to enhance immunogenicity. Effective

polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals response vary depending on site of inoculations and doses, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, *et al.* *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, *et al.*, Chap. 19 in: *Handbook of Experimental Immunology* D. Wier (ed) Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12  $\mu$ M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980).

Antibody preparations prepared according to either of the above protocols are useful in a variety of contexts. In particular, the antibodies may be used in immunoaffinity chromatography techniques such as those described below to facilitate large scale isolation, purification, or enrichment of the proteins or polypeptides encoded by EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or for the isolation, purification or enrichment of EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides or fragments of positional segments of EST-related polypeptides.

In the case of secreted proteins, the antibodies may be used for the isolation, purification, or enrichment of the full protein (*i.e.* the mature protein and the signal peptide), the mature protein (*i.e.* the protein generated by cleavage of the signal peptide), or the signal peptide are operably linked to promoters and introduced into cells as described above.

Additionally, the antibodies may be used in immunoaffinity chromatography techniques such as those described below to isolate, purify, or enrich polypeptides which have been linked to the proteins or polypeptides encoded by EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or to isolate, purify, or enrich EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides or fragments of positional segments of EST-related polypeptides.

The antibodies may also be used to determine the cellular localization of polypeptides encoded by the proteins or polypeptides encoded by EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or the cellular

localization of EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides or fragments of positional segments of EST-related polypeptides.

In addition, the antibodies may also be used to determine the cellular localization of polypeptides which have been linked to the proteins or polypeptides encoded by EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or polypeptides which have been linked to EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides or fragments of positional segments of EST-related polypeptides .

The antibodies may also be used in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they may also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample or to identify the type of tissue present in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

## 15 VI. Use of 5'ESTs or Consensus Contigated 5' ESTs or Sequences Obtainable Therefrom or Portions Thereof as Reagents

The EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be used as reagents in isolation procedures, diagnostic assays, and forensic procedures. For example, sequences from the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids, may be detectably labeled and used as probes to isolate other sequences capable of hybridizing to them. In addition, the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be used to design PCR primers to be used in isolation, diagnostic, or forensic procedures.

25

### 1. Use of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids in isolation, diagnostic and forensic procedures

#### EXAMPLE 36

##### Preparation of PCR Primers and Amplification of DNA

30 The EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be used to prepare PCR primers for a variety of applications, including isolation procedures for cloning nucleic acids capable of hybridizing to such sequences, diagnostic techniques and forensic techniques. In some embodiments, the PCR primers at least 10, 15, 18, 20, 23, 25, 28, 30, 40, or 50 nucleotides in length. In some embodiments, the PCR primers may be more than 30 bases in length. It is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see Molecular Cloning to

35

Genetic Engineering White, B.A. Ed. in *Methods in Molecular Biology* 67: Humana Press, Totowa 1997.

In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites.

10

### EXAMPLE 37

#### Use of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids as probes

Probes derived from EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be labeled with detectable labels familiar to those skilled in the art, including radioisotopes and non-radioactive labels, to provide a detectable probe. The detectable probe may be single stranded or double stranded and may be made using techniques known in the art, including *in vitro* transcription, nick translation, or kinase reactions. A nucleic acid sample containing a sequence capable of hybridizing to the labeled probe is contacted with the labeled probe. If the nucleic acid in the sample is double stranded, it may be denatured prior to contacting the probe. In some applications, the nucleic acid sample may be immobilized on a surface such as a nitrocellulose or nylon membrane. The nucleic acid sample may comprise nucleic acids obtained from a variety of sources, including genomic DNA, cDNA libraries, RNA, or tissue samples.

Procedures used to detect the presence of nucleic acids capable of hybridizing to the detectable probe include well known techniques such as Southern blotting, Northern blotting, dot blotting, colony hybridization, and plaque hybridization. In some applications, the nucleic acid capable of hybridizing to the labeled probe may be cloned into vectors such as expression vectors, sequencing vectors, or *in vitro* transcription vectors to facilitate the characterization and expression of the hybridizing nucleic acids in the sample. For example, such techniques may be used to isolate and clone sequences in a genomic library or cDNA library which are capable of hybridizing to the detectable probe as described in Example 20 above.

PCR primers made as described in Example 36 above may be used in forensic analyses, such as the DNA fingerprinting techniques described in Examples 38-42 below. Such analyses may utilize detectable probes or primers based on the sequences of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids.

35

### EXAMPLE 38

#### Forensic Matching by DNA Sequencing



In one exemplary method, DNA samples are isolated from forensic specimens of, for example, hair, semen, blood or skin cells by conventional methods. A panel of PCR primers based on a number of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids is then utilized in accordance with Example 36 to  
5 amplify DNA of approximately 100-200 bases in length from the forensic specimen. Corresponding sequences are obtained from a test subject. Each of these identification DNAs is then sequenced using standard techniques, and a simple database comparison determines the differences, if any, between the sequences from the subject and those from the sample. Statistically significant differences between the suspect's DNA sequences and those from the sample conclusively prove a lack of identity. This lack of  
10 identity can be proven, for example, with only one sequence. Identity, on the other hand, should be demonstrated with a large number of sequences, all matching. Preferably, a minimum of 50 statistically identical sequences of 100 bases in length are used to prove identity between the suspect and the sample.

#### EXAMPLE 39

##### 15 Positive Identification by DNA Sequencing

The technique outlined in the previous example may also be used on a larger scale to provide a unique fingerprint-type identification of any individual. In this technique, primers are prepared from a large number of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. Preferably, 20 to 50 different primers are  
20 used. These primers are used to obtain a corresponding number of PCR-generated DNA segments from the individual in question in accordance with Example 34. Each of these DNA segments is sequenced, using the methods set forth in Example 36. The database of sequences generated through this procedure uniquely identifies the individual from whom the sequences were obtained. The same panel of primers may then be used at any later time to absolutely correlate tissue or other biological specimen with that  
25 individual.

#### EXAMPLE 40

##### Southern Blot Forensic Identification

The procedure of Example 38 is repeated to obtain a panel of at least 10 amplified sequences  
30 from an individual and a specimen. Preferably, the panel contains at least 50 amplified sequences. More preferably, the panel contains 100 amplified sequences. In some embodiments, the panel contains 200 amplified sequences. This PCR-generated DNA is then digested with one or a combination of, preferably, four base specific restriction enzymes. Such enzymes are commercially available and known to those of skill in the art. After digestion, the resultant gene fragments are size separated in multiple  
35 duplicate wells on an agarose gel and transferred to nitrocellulose using Southern blotting techniques well known to those with skill in the art. For a review of Southern blotting see Davis *et al.* (Basic Methods in Molecular Biology, 1986, Elsevier Press. pp 62-65).

A panel of probes based on the sequences of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are radioactively or colorimetrically labeled using methods known in the art, such as nick translation or end labeling, and hybridized to the Southern blot using techniques known in the art (Davis *et al.*, supra).

- 5 Preferably, the probe is at least 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 or 500 nucleotides in length. Preferably, the probes are at least 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 or 500 nucleotides in length. In some embodiments, the probes are oligonucleotides which are 40 nucleotides in length or less.

- Preferably, at least 5 to 10 of these labeled probes are used, and more preferably at least about  
10 20 or 30 are used to provide a unique pattern. The resultant bands appearing from the hybridization of a large sample of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids will be a unique identifier. Since the restriction enzyme cleavage will be different for every individual, the band pattern on the Southern blot will also be unique. Increasing the number of probes will provide a statistically higher level of confidence in the  
15 identification since there will be an increased number of sets of bands used for identification.

#### EXAMPLE 41

##### Dot Blot Identification Procedure

- Another technique for identifying individuals using the EST-related nucleic acids, positional  
20 segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids disclosed herein utilizes a dot blot hybridization technique.

- Genomic DNA is isolated from nuclei of subject to be identified. Probes are prepared that correspond to at least 10, preferably 50 sequences from the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids.  
25 The probes are used to hybridize to the genomic DNA through conditions known to those in the art. The oligonucleotides are end labeled with  $P^{32}$  using polynucleotide kinase (Pharmacia). Dot Blots are created by spotting the genomic DNA onto nitrocellulose or the like using a vacuum dot blot manifold (BioRad, Richmond California). The nitrocellulose filter containing the genomic sequences is baked or UV linked to the filter, prehybridized and hybridized with labeled probe using techniques known in the art (Davis *et al.*, supra). The  $^{32}P$  labeled DNA fragments are sequentially hybridized with successively stringent  
30 conditions to detect minimal differences between the 30 bp sequence and the DNA.

Tetramethylammonium chloride is useful for identifying clones containing small numbers of nucleotide mismatches (Wood *et al.*, *Proc. Natl. Acad. Sci. USA* 82(6):1585-1588 (1985)). A unique pattern of dots distinguishes one individual from another individual.

- 35 EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids can be used as probes in the following alternative

fingerprinting technique. In some embodiments, the probes are oligonucleotides which are 40 nucleotides in length or less.

Preferably, a plurality of probes having sequences from different EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are used in the alternative fingerprinting technique. Example 42 below provides a representative alternative fingerprinting procedure in which the probes are derived from EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids.

10

#### EXAMPLE 42

##### Alternative "Fingerprint" Identification Technique

Oligonucleotides are prepared from a large number, e.g. 50, 100, or 200, EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids using commercially available oligonucleotide services such as Genset, Paris, France. Preferably, the oligonucleotides are at least 10, 15, 18, 20, 23, 25, 28, or 30 nucleotides in length. However, in some embodiments, the oligonucleotides may be more than 40, 50, 60 or 70 nucleotides in length.

Cell samples from the test subject are processed for DNA using techniques well known to those with skill in the art. The nucleic acid is digested with restriction enzymes such as EcoRI and XbaI. Following digestion, samples are applied to wells for electrophoresis. The procedure, as known in the art, may be modified to accommodate polyacrylamide electrophoresis, however in this example, samples containing 5 ug of DNA are loaded into wells and separated on 0.8% agarose gels. The gels are transferred onto nitrocellulose using standard Southern blotting techniques.

10 ng of each of the oligonucleotides are pooled and end-labeled with P<sup>32</sup>. The nitrocellulose is prehybridized with blocking solution and hybridized with the labeled probes. Following hybridization and washing, the nitrocellulose filter is exposed to X-Omat AR X-ray film. The resulting hybridization pattern will be unique for each individual.

It is additionally contemplated within this example that the number of probe sequences used can be varied for additional accuracy or clarity.

In addition to their applications in forensics and identification, EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be mapped to their chromosomal locations. Example 41 below describes radiation hybrid (RH) mapping of human chromosomal regions using EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. Example 42 below describes a representative procedure for mapping EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids to their locations on human chromosomes. Example 43 below describes mapping of

EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids on metaphase chromosomes by Fluorescence In Situ Hybridization (FISH).

- 5 2. Use of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids in Chromosome Mapping

#### EXAMPLE 43

Radiation hybrid mapping of EST-related nucleic acids, positional segments of

EST-related nucleic acids or fragments of positional segments of

10 EST-related nucleic acids to the human genome

Radiation hybrid (RH) mapping is a somatic cell genetic approach that can be used for high resolution mapping of the human genome. In this approach, cell lines containing one or more human chromosomes are lethally irradiated, breaking each chromosome into fragments whose size depends on the radiation dose. These fragments are rescued by fusion with cultured rodent cells, yielding subclones  
15 containing different portions of the human genome. This technique is described by Benham *et al.* (*Genomics* 4:509-517, 1989) and Cox *et al.*, (*Science* 250:245-250, 1990). The random and independent nature of the subclones permits efficient mapping of any human genome marker. Human DNA isolated from a panel of 80-100 cell lines provides a mapping reagent for ordering EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related  
20 nucleic acids. In this approach, the frequency of breakage between markers is used to measure distance, allowing construction of fine resolution maps as has been done using conventional ESTs (Schuler *et al.*, *Science* 274:540-546, 1996).

RH mapping has been used to generate a high-resolution whole genome radiation hybrid map of human chromosome 17q22-q25.3 across the genes for growth hormone (GH) and thymidine kinase (TK)  
25 (Foster *et al.*, *Genomics* 33:185-192, 1996), the region surrounding the Gorlin syndrome gene (Obermayr *et al.*, *Eur. J. Hum. Genet.* 4:242-245, 1996), 60 loci covering the entire short arm of chromosome 12 (Raeymaekers *et al.*, *Genomics* 29:170-178, 1995), the region of human chromosome 22 containing the neurofibromatosis type 2 locus (Frazer *et al.*, *Genomics* 14:574-584, 1992) and 13 loci on the long arm of chromosome 5 (Warrington *et al.*, *Genomics* 11:701-708, 1991).

30

#### EXAMPLE 44

Mapping of EST-related nucleic acids, positional segments of

EST-related nucleic acids or fragments of positional segments of

EST-related nucleic acids to Human Chromosomes using PCR techniques

35 EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be assigned to human chromosomes using PCR based methodologies. In such approaches, oligonucleotide primer pairs are designed from EST-related

nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids to minimize the chance of amplifying through an intron. Preferably, the oligonucleotide primers are 18-23 bp in length and are designed for PCR amplification. The creation of PCR primers from known sequences is well known to those with skill in the art. For a review of PCR technology see Erlich, in PCR Technology; Principles and Applications for DNA Amplification. 1992. W.H. Freeman and Co., New York.

The primers are used in polymerase chain reactions (PCR) to amplify templates from total human genomic DNA. PCR conditions are as follows: 60 ng of genomic DNA is used as a template for PCR with 80 ng of each oligonucleotide primer, 0.6 unit of Taq polymerase, and 1  $\mu$ Cu of a  $^{32}$ P-labeled deoxycytidine triphosphate. The PCR is performed in a microplate thermocycler (Technique) under the following conditions: 30 cycles of 94°C, 1.4 min; 55°C, 2 min; and 72°C, 2 min; with a final extension at 72°C for 10 min. The amplified products are analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. If the length of the resulting PCR product is identical to the distance between the ends of the primer sequences in the 5'EST from which the primers are derived, then the PCR reaction is repeated with DNA templates from two panels of human-rodent somatic cell hybrids, BIOS PCRable DNA (BIOS Corporation) and NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel Number 1 (NIGMS, Camden, NJ).

PCR is used to screen a series of somatic cell hybrid cell lines containing defined sets of human chromosomes for the presence of a given 5'EST. DNA is isolated from the somatic hybrids and used as starting templates for PCR reactions using the primer pairs from the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. Only those somatic cell hybrids with chromosomes containing the human gene corresponding to the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids will yield an amplified fragment. The 5'ESTs are assigned to a chromosome by analysis of the segregation pattern of PCR products from the somatic hybrid DNA templates. The single human chromosome present in all cell hybrids that give rise to an amplified fragment is the chromosome containing that EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. For a review of techniques and analysis of results from somatic cell gene mapping experiments. (See Ledbetter *et al.*, Genomics 6:475-481 (1990)).

Alternatively, the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be mapped to individual chromosomes using FISH as described in Example 45 below.

#### EXAMPLE 45

Mapping of EST-related nucleic acids, positional segments of  
EST-related nucleic acids or fragments of positional segments of

EST-related nucleic acids to Chromosomes Using  
Fluorescence In Situ Hybridization

Fluorescence in situ hybridization allows the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids to be mapped to a particular location on a given chromosome. The chromosomes to be used for fluorescence in situ hybridization techniques may be obtained from a variety of sources including cell cultures, tissues, or whole blood.

In a preferred embodiment, chromosomal localization of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are obtained by FISH as described by Cherif *et al.* (*Proc. Natl. Acad. Sci. U.S.A.*, 87:6639-6643, 1990). Metaphase chromosomes are prepared from phytohemagglutinin (PHA)-stimulated blood cell donors. PHA-stimulated lymphocytes from healthy males are cultured for 72 h in RPMI-1640 medium. For synchronization, methotrexate (10  $\mu$ M) is added for 17 h, followed by addition of 5-bromodeoxyuridine (5-BrdU, 0.1 mM) for 6 h. Colcemid (1  $\mu$ g/ml) is added for the last 15 min before harvesting the cells. Cells are collected, washed in RPMI, incubated with a hypotonic solution of KCl (75 mM) at 37°C for 15 min and fixed in three changes of methanol:acetic acid (3:1). The cell suspension is dropped onto a glass slide and air dried. The EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids is labeled with biotin-16 dUTP by nick translation according to the manufacturer's instructions (Bethesda Research Laboratories, Bethesda, MD), purified using a Sephadex G-50 column (Pharmacia, Upsala, Sweden) and precipitated. Just prior to hybridization, the DNA pellet is dissolved in hybridization buffer (50% formamide, 2 X SSC, 10% dextran sulfate, 1 mg/ml sonicated salmon sperm DNA, pH 7) and the probe is denatured at 70°C for 5-10 min.

Slides kept at -20°C are treated for 1 h at 37°C with RNase A (100  $\mu$ g/ml), rinsed three times in 2 X SSC and dehydrated in an ethanol series. Chromosome preparations are denatured in 70% formamide, 2 X SSC for 2 min at 70°C, then dehydrated at 4°C. The slides are treated with proteinase K (10  $\mu$ g/100 ml in 20 mM Tris-HCl, 2 mM CaCl<sub>2</sub>) at 37°C for 8 min and dehydrated. The hybridization mixture containing the probe is placed on the slide, covered with a coverslip, sealed with rubber cement and incubated overnight in a humid chamber at 37°C. After hybridization and post-hybridization washes, the biotinylated probe is detected by avidin-FITC and amplified with additional layers of biotinylated goat anti-avidin and avidin-FITC. For chromosomal localization, fluorescent R-bands are obtained as previously described (Cherif *et al.*, *supra.*). The slides are observed under a LEICA fluorescence microscope (DMRXA). Chromosomes are counterstained with propidium iodide and the fluorescent signal of the probe appears as two symmetrical yellow-green spots on both chromatids of the fluorescent R-band chromosome (red). Thus, a particular EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be localized to a particular cytogenetic R-band on a given chromosome.

Once the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids have been assigned to particular chromosomes using the techniques described in Examples 42-44 above, they may be utilized to construct a high resolution map of the chromosomes on which they are located or to identify the chromosomes in a sample.

#### EXAMPLE 46

##### Use of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids to Construct or Expand Chromosome Maps

Chromosome mapping involves assigning a given unique sequence to a particular chromosome as described above. Once the unique sequence has been mapped to a given chromosome, it is ordered relative to other unique sequences located on the same chromosome. One approach to chromosome mapping utilizes a series of yeast artificial chromosomes (YACs) bearing several thousand long inserts derived from the chromosomes of the organism from which the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are obtained. This approach is described in Ramaiah Nagaraja *et al.*, *Genome Research* 7:210-222, March 1997. Briefly, in this approach each chromosome is broken into overlapping pieces which are inserted into the YAC vector. The YAC inserts are screened using PCR or other methods to determine whether they include the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids whose position is to be determined. Once an insert has been found which includes the 5'EST, the insert can be analyzed by PCR or other methods to determine whether the insert also contains other sequences known to be on the chromosome or in the region from which the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids was derived. This process can be repeated for each insert in the YAC library to determine the location of each of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids relative to one another and to other known chromosomal markers. In this way, a high resolution map of the distribution of numerous unique markers along each of the organisms chromosomes may be obtained.

As described in Example 47 below EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may also be used to identify genes associated with a particular phenotype, such as hereditary disease or drug response.

##### 3. Use of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids Gene Identification

#### EXAMPLE 47

##### Identification of genes associated with hereditary diseases or drug response

This example illustrates an approach useful for the association of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids with particular phenotypic characteristics. In this example, a particular EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids is used as a test probe to associate that EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids with a particular phenotypic characteristic.

EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are mapped to a particular location on a human chromosome using techniques such as those described in Examples 41 and 42 or other techniques known in the art. A search of Mendelian Inheritance in Man (V. McKusick, *Mendelian Inheritance in Man* (available on line through Johns Hopkins University Welch Medical Library) reveals the region of the human chromosome which contains the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids to be a very gene rich region containing several known genes and several diseases or phenotypes for which genes have not been identified. The gene corresponding to this EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids thus becomes an immediate candidate for each of these genetic diseases.

Cells from patients with these diseases or phenotypes are isolated and expanded in culture. PCR primers from the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are used to screen genomic DNA, mRNA or cDNA obtained from the patients. EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids that are not amplified in the patients can be positively associated with a particular disease by further analysis. Alternatively, the PCR analysis may yield fragments of different lengths when the samples are derived from an individual having the phenotype associated with the disease than when the sample is derived from a healthy individual, indicating that the gene containing the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be responsible for the genetic disease.

## **VII. Use of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids to Construct Vectors**

The present EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may also be used to construct secretion vectors capable of directing the secretion of the proteins encoded by genes therein. Such secretion vectors may facilitate the purification or enrichment of the proteins encoded by genes inserted therein by



reducing the number of background proteins from which the desired protein must be purified or enriched. Exemplary secretion vectors are described in Example 48 below.

### 1. Construction of secretion vectors

### EXAMPLE 48

## 5 Construction of Secretion Vectors

The secretion vectors of the present invention include a promoter capable of directing gene expression in the host cell, tissue, or organism of interest. Such promoters include the Rous Sarcoma Virus promoter, the SV40 promoter, the human cytomegalovirus promoter, and other promoters familiar to those skilled in the art.

10           A signal sequence from one of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids is operably linked to the promoter such that the mRNA transcribed from the promoter will direct the translation of the signal peptide. Preferably, the signal sequence is from one of the nucleic acids of SEQ ID NOs.24-811. The host cell, tissue, or organism may be any cell, tissue, or organism which recognizes the signal

15           peptide encoded by the signal sequence in the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. Suitable hosts include mammalian cells, tissues or organisms, avian cells, tissues, or organisms, insect cells, tissues or organisms, or yeast.

In addition, the secretion vector contains cloning sites for inserting genes encoding the proteins which are to be secreted. The cloning sites facilitate the cloning of the insert gene in frame with the signal sequence such that a fusion protein in which the signal peptide is fused to the protein encoded by the inserted gene is expressed from the mRNA transcribed from the promoter. The signal peptide directs the extracellular secretion of the fusion protein.

The secretion vector may be DNA or RNA and may integrate into the chromosome of the host, be stably maintained as an extrachromosomal replicon in the host, be an artificial chromosome, or be transiently present in the host. Preferably, the secretion vector is maintained in multiple copies in each host cell. As used herein, multiple copies means at least 2, 5, 10, 20, 25, 50 or more than 50 copies per cell. In some embodiments, the multiple copies are maintained extrachromosomally. In other embodiments, the multiple copies result from amplification of a chromosomal sequence.

30 Many nucleic acid backbones suitable for use as secretion vectors are known to those skilled in the art, including retroviral vectors, SV40 vectors, Bovine Papilloma Virus vectors, yeast integrating plasmids, yeast episomal plasmids, yeast artificial chromosomes, human artificial chromosomes, P element vectors, baculovirus vectors, or bacterial plasmids capable of being transiently introduced into the host.

35 The secretion vector may also contain a polyA signal such that the polyA signal is located downstream of the gene inserted into the secretion vector.

After the gene encoding the protein for which secretion is desired is inserted into the secretion vector, the secretion vector is introduced into the host cell, tissue, or organism using calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection, viral particles or as naked DNA. The protein encoded by the inserted gene is then purified or enriched from the supernatant using conventional techniques such as ammonium sulfate precipitation, immunoprecipitation, immunoaffinitychromatography, size exclusion chromatography, ion exchange chromatography, and HPLC. Alternatively, the secreted protein may be in a sufficiently enriched or pure state in the supernatant or growth media of the host to permit it to be used for its intended purpose without further enrichment.

The signal sequences may also be inserted into vectors designed for gene therapy. In such vectors, the signal sequence is operably linked to a promoter such that mRNA transcribed from the promoter encodes the signal peptide. A cloning site is located downstream of the signal sequence such that a gene encoding a protein whose secretion is desired may readily be inserted into the vector and fused to the signal sequence. The vector is introduced into an appropriate host cell. The protein expressed from the promoter is secreted extracellularly, thereby producing a therapeutic effect.

#### EXAMPLE 49

##### Fusion Vectors

The EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be used to construct fusion vectors for the expression of chimeric polypeptides. The chimeric polypeptides comprise a first polypeptide portion and a second polypeptide portion. In the fusion vectors of the present invention, nucleic acids encoding the first polypeptide portion and the second polypeptide portion are joined in frame with one another so as to generate a nucleic acid encoding the chimeric polypeptide. The nucleic acid encoding the chimeric polypeptide is operably linked to a promoter which directs the expression of an mRNA encoding the chimeric polypeptide. The promoter may be in any of the expression vectors described herein including those described in Examples 21 and 48.

Preferably, the fusion vector is maintained in multiple copies in each host cell. In some embodiments, the multiple copies are maintained extrachromosomally. In other embodiments, the multiple copies result from amplification of a chromosomal sequence.

The first polypeptide portion may comprise any of the polypeptides encoded by the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. In some embodiments, the first polypeptide portion may be one of the EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides, or fragments of positional segments of EST-related polypeptides.

The second polypeptide portion may comprise any polypeptide of interest. In some embodiments, the second polypeptide portion may comprise a polypeptide having a detectable

- enzymatic activity such as green fluorescent protein or  $\beta$  galactosidase. Chimeric polypeptides in which the second polypeptide portion comprises a detectable polypeptide may be used to determine the intracellular localization of the first polypeptide portion. In such procedures, the fusion vector encoding the chimeric polypeptide is introduced into a host cell under conditions which facilitate the expression of the chimeric polypeptide. Where appropriate, the cells are treated with a detection reagent which is visible under the microscope following a catalytic reaction with the detectable polypeptide and the cellular location of the detection reagent is determined. For example, if the polypeptide having a detectable enzymatic activity is  $\beta$  galactosidase, the cells may be treated with Xgal. Alternatively, where the detectable polypeptide is directly detectable without the addition of a detection reagent, the intracellular location of the chimeric polypeptide is determined by performing microscopy under conditions in which the detectable polypeptide is visible. For example, if the detectable polypeptide is green fluorescent protein or a modified version thereof, microscopy is performed by exposing the host cells to light having an appropriate wavelength to cause the green fluorescent protein or modified version thereof to fluoresce.
- Alternatively, the second polypeptide portion may comprise a polypeptide whose isolation, purification, or enrichment is desired. In such embodiments, the isolation, purification, or enrichment of the second polypeptide portion may be achieved by performing the immunoaffinity chromatography procedures described below using an immunoaffinity column having an antibody directed against the first polypeptide portion coupled thereto.
- The proteins encoded by the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or the EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides, or fragments of positional segments of EST-related polypeptides may also be used to generate antibodies as explained herein in order to identify the tissue type or cell species from which a sample is derived as described in Example 50.

### EXAMPLE 50

#### Identification of Tissue Types or Cell Species by Means of

##### Labeled Tissue Specific Antibodies

- Identification of specific tissues is accomplished by the visualization of tissue specific antigens by means of antibody preparations as described herein which are conjugated, directly or indirectly to a detectable marker. Selected labeled antibody species bind to their specific antigen binding partner in tissue sections, cell suspensions, or in extracts of soluble proteins from a tissue sample to provide a pattern for qualitative or semi-qualitative interpretation.
- Antisera for these procedures must have a potency exceeding that of the native preparation, and for that reason, antibodies are concentrated to a mg/ml level by isolation of the gamma globulin fraction, for example, by ion-exchange chromatography or by ammonium sulfate fractionation. Also, to provide

the most specific antisera, unwanted antibodies, for example to common proteins, must be removed from the gamma globulin fraction, for example by means of insoluble immunoabsorbents, before the antibodies are labeled with the marker. Either monoclonal or heterologous antisera is suitable for either procedure.

## 5 1. Immunohistochemical Techniques

Purified, high-titer antibodies, prepared as described above, are conjugated to a detectable marker, as described, for example, by Fudenberg, H., Chap. 26 in: *Basic 503 Clinical Immunology*, 3<sup>rd</sup> Ed. Lange, Los Altos, California (1980) or Rose, *et al.*, Chap. 12 in: *Methods in Immunodiagnosis*, 2d Ed. John Wiley and Sons, New York (1980).

10 A fluorescent marker, either fluorescein or rhodamine, is preferred, but antibodies can also be labeled with an enzyme that supports a color producing reaction with a substrate, such as horseradish peroxidase. Markers can be added to tissue-bound antibody in a second step, as described below. Alternatively, the specific antitissue antibodies can be labeled with ferritin or other electron dense particles, and localization of the ferritin coupled antigen-antibody complexes achieved by means of an  
15 electron microscope. In yet another approach, the antibodies are radiolabeled, with, for example <sup>125</sup>I, and detected by overlaying the antibody treated preparation with photographic emulsion.

Preparations to carry out the procedures can comprise monoclonal or polyclonal antibodies to a single protein or peptide identified as specific to a tissue type, for example, brain tissue, or antibody preparations to several antigenically distinct tissue specific antigens can be used in panels, independently  
20 or in mixtures, as required.

Tissue sections and cell suspensions are prepared for immunohistochemical examination according to common histological techniques. Multiple cryostat sections (about 4 μm, unfixed) of the unknown tissue and known control, are mounted and each slide covered with different dilutions of the antibody preparation. Sections of known and unknown tissues should also be treated with preparations  
25 to provide a positive control, a negative control, for example, pre-immune sera, and a control for non-specific staining, for example, buffer.

Treated sections are incubated in a humid chamber for 30 min at room temperature, rinsed, then washed in buffer for 30-45 min. Excess fluid is blotted away, and the marker developed.

If the tissue specific antibody was not labeled in the first incubation, it can be labeled at this time  
30 in a second antibody-antibody reaction, for example, by adding fluorescein- or enzyme-conjugated antibody against the immunoglobulin class of the antiserum-producing species, for example, fluorescein labeled antibody to mouse IgG. Such labeled sera are commercially available.

The antigen found in the tissues by the above procedure can be quantified by measuring the intensity of color or fluorescence on the tissue section, and calibrating that signal using appropriate  
35 standards.

## 2. Identification of Tissue Specific Soluble Proteins

The visualization of tissue specific proteins and identification of unknown tissues from that procedure is carried out using the labeled antibody reagents and detection strategy as described for immunohistochemistry; however the sample is prepared according to an electrophoretic technique to distribute the proteins extracted from the tissue in an orderly array on the basis of molecular weight for  
5 detection.

A tissue sample is homogenized using a Virtis apparatus; cell suspensions are disrupted by Dounce homogenization or osmotic lysis, using detergents in either case as required to disrupt cell membranes, as is the practice in the art. Insoluble cell components such as nuclei, microsomes, and membrane fragments are removed by ultracentrifugation, and the soluble protein-containing fraction  
10 concentrated if necessary and reserved for analysis.

A sample of the soluble protein solution is resolved into individual protein species by conventional SDS polyacrylamide electrophoresis as described, for example, by Davis, L. *et al.*, Section 19-2 in: *Basic Methods in Molecular Biology* (P. Leder, ed), Elsevier, New York (1986), using a range of amounts of polyacrylamide in a set of gels to resolve the entire molecular weight range of proteins to  
15 be detected in the sample. A size marker is run in parallel for purposes of estimating molecular weights of the constituent proteins. Sample size for analysis is a convenient volume of from 5 to 55  $\mu$ l, and containing from about 1 to 100  $\mu$ g protein. An aliquot of each of the resolved proteins is transferred by blotting to a nitrocellulose filter paper, a process that maintains the pattern of resolution. Multiple copies are prepared. The procedure, known as Western Blot Analysis, is well described in Davis, L. *et al.*,  
20 *supra* Section 19-3. One set of nitrocellulose blots is stained with Coomassie Blue dye to visualize the entire set of proteins for comparison with the antibody bound proteins. The remaining nitrocellulose filters are then incubated with a solution of one or more specific antisera to tissue specific proteins prepared as described in Examples 20 and 33. In this procedure, as in procedure A above, appropriate positive and negative sample and reagent controls are run.

In either procedure described above a detectable label can be attached to the primary tissue antigen-primary antibody complex according to various strategies and permutations thereof. In a straightforward approach, the primary specific antibody can be labeled; alternatively, the unlabeled complex can be bound by a labeled secondary anti-IgG antibody. In other approaches, either the primary or secondary antibody is conjugated to a biotin molecule, which can, in a subsequent step, bind an avidin  
30 conjugated marker. According to yet another strategy, enzyme labeled or radioactive protein A, which has the property of binding to any IgG, is bound in a final step to either the primary or secondary antibody.

#### EXAMPLE 51

##### 35 Immunohistochemical Localization of Polypeptides

The antibodies prepared as described herein above may be utilized to determine the cellular location of a polypeptide. The polypeptide may be any of the polypeptides encoded by EST-related

nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or the polypeptide may be one of the EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides, or fragments of positional segments of EST-related polypeptides. In some embodiments, the polypeptide may be a chimeric  
5 polypeptide such as those encoded by the fusion vectors of Example 49.

Cells expressing the polypeptide to be localized are applied to a microscope slide and fixed using any of the procedures typically employed in immunohistochemical localization techniques, including the methods described in *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc. 1997. Following a washing step, the cells are contacted with the antibody. In some embodiments, the  
10 antibody is conjugated to a detectable marker as described above to facilitate detection. Alternatively, in some embodiments, after the cells have been contacted with an antibody to the polypeptide to be localized, a secondary antibody which has been conjugated to a detectable marker is placed in contact with the antibody against the polypeptide to be localized.

Thereafter, microscopy is performed under conditions suitable for visualizing the cellular  
15 location of the polypeptide.

The visualization of tissue specific antigen binding at levels above those seen in control tissues to one or more tissue specific antibodies, directed against the polypeptides encoded by EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or antibodies against the EST-related polypeptides, fragments of EST-related  
20 polypeptides, positional segments of EST-related polypeptides, or fragments of positional segments of EST-related polypeptides, can identify tissues of unknown origin, for example, forensic samples, or differentiated tumor tissue that has metastasized to foreign bodily sites.

The antibodies described herein may also be used in the immunoaffinity chromatography techniques described below to isolate, purify or enrich the polypeptides encoded by the EST-related  
25 nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or to isolate, purify or enrich EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides, or fragments of positional segments of EST-related polypeptides. The immunoaffinity chromatography techniques described below may also be used to isolate, purify or enrich polypeptides which have been linked to the  
30 polypeptides encoded by the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or to isolate, purify or enrich polypeptides which have been linked to EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides, or fragments of positional segments of EST-related polypeptides.

**EXAMPLE 52**Immunoaffinity Chromatography

Antibodies prepared as described above are coupled to a support. Preferably, the antibodies are monoclonal antibodies, but polyclonal antibodies may also be used. The support may be any of those typically employed in immunoaffinity chromatography, including Sepharose CL-4B (Pharmacia, Piscataway, NJ), Sepharose CL-2B (Pharmacia, Piscataway, NJ), Affi-gel 10 (Biorad, Richmond, CA), or glass beads.

The antibodies may be coupled to the support using any of the coupling reagents typically used in immunoaffinity chromatography, including cyanogen bromide. After coupling the antibody to the support, the support is contacted with a sample which contains a target polypeptide whose isolation, purification or enrichment is desired. The target polypeptide may be a polypeptide encoded by the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or the target polypeptide may be one of the EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides, or fragments of positional segments of EST-related polypeptides. The target polypeptides may also be polypeptides which have been linked to the polypeptides encoded by the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or the target polypeptides may be polypeptides which have been linked to EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides, or fragments of positional segments of EST-related polypeptides using the fusion vectors described above.

Preferably, the sample is placed in contact with the support for a sufficient amount of time and under appropriate conditions to allow at least 50% of the target polypeptide to specifically bind to the antibody coupled to the support.

Thereafter, the support is washed with an appropriate wash solution to remove polypeptides which have non-specifically adhered to the support. The wash solution may be any of those typically employed in immunoaffinity chromatography, including PBS, Tris-lithium chloride buffer (0.1M lysine base and 0.5M lithium chloride, pH 8.0), Tris-hydrochloride buffer (0.05M Tris-hydrochloride, pH 8.0), or Tris/Triton/NaCl buffer (50mM Tris.cl, pH 8.0 or 9.0, 0.1% Triton X-100, and 0.5MNaCl).

After washing, the specifically bound target polypeptide is eluted from the support using the high pH or low pH elution solutions typically employed in immunoaffinity chromatography. In particular, the elution solutions may contain an eluant such as triethanolamine, diethylamine, calcium chloride, sodium thiocyanate, potassium bromide, acetic acid, or glycine. In some embodiments, the elution solution may also contain a detergent such as Triton X-100 or octyl- $\beta$ -D-glucoside.

The EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may also be used to clone sequences located upstream of the 5'ESTs which are capable of regulating gene expression, including promoter sequences, enhancer

sequences, and other upstream sequences which influence transcription or translation levels. Once identified and cloned, these upstream regulatory sequences may be used in expression vectors designed to direct the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative fashion. Example 51 describes a method for cloning sequences upstream of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids.

## 2. Identification of upstream sequences with promoting or regulatory activities

### **EXAMPLE 53**

#### 10    Use of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids to Clone Upstream Sequences from Genomic DNA

Sequences derived from EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be used to isolate the promoters of the corresponding genes using chromosome walking techniques. In one chromosome walking technique, which utilizes the GenomeWalker™ kit available from Clontech, five complete genomic DNA samples are each digested with a different restriction enzyme which has a 6 base recognition site and leaves a blunt end. Following digestion, oligonucleotide adapters are ligated to each end of the resulting genomic DNA fragments.

For each of the five genomic DNA libraries, a first PCR reaction is performed according to the manufacturer's instructions using an outer adapter primer provided in the kit and an outer gene specific primer. The gene specific primer should be selected to be specific for 5' EST of interest and should have a melting temperature, length, and location in the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids which is consistent with its use in PCR reactions. Each first PCR reaction contains 5ng of genomic DNA, 5 µl of 10X Tth reaction buffer, 0.2 mM of each dNTP, 0.2 µM each of outer adapter primer and outer gene specific primer, 1.1 mM of Mg(OAc)<sub>2</sub>, and 1 µl of the Tth polymerase 50X mix in a total volume of 50 µl. The reaction cycle for the first PCR reaction is as follows: 1 min at 94°C / 2 sec at 94°C, 3 min at 72°C (7 cycles) / 2 sec at 94°C, 3 min at 67°C (32 cycles) / 5 min at 67°C.

The product of the first PCR reaction is diluted and used as a template for a second PCR reaction according to the manufacturer's instructions using a pair of nested primers which are located internally on the amplicon resulting from the first PCR reaction. For example, 5 µl of the reaction product of the first PCR reaction mixture may be diluted 180 times. Reactions are made in a 50 µl volume having a composition identical to that of the first PCR reaction except the nested primers are used. The first nested primer is specific for the adapter, and is provided with the GenomeWalker™ kit. The second nested primer is specific for the particular EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids for which the promoter is to be cloned and should have a melting temperature, length, and location in



the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids which is consistent with its use in PCR reactions. The reaction parameters of the second PCR reaction are as follows: 1 min at 94°C / 2 sec at 94°C, 3 min at 72°C (6 cycles) / 2 sec at 94°C, 3 min at 67°C (25 cycles) / 5 min at 67°C. The product of the second PCR reaction is purified, cloned, and sequenced using standard techniques.

Alternatively, two or more human genomic DNA libraries can be constructed by using two or more restriction enzymes. The digested genomic DNA is cloned into vectors which can be converted into single stranded, circular, or linear DNA. A biotinylated oligonucleotide comprising at least 10, 12, 15, 18, 20, 23, 25, 27, 30, 35, 40, or 50 nucleotides from the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids sequence is hybridized to the single stranded DNA. Hybrids between the biotinylated oligonucleotide and the single stranded DNA containing the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are isolated as described above. Thereafter, the single stranded DNA containing the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids is released from the beads and converted into double stranded DNA using a primer specific for the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or a primer corresponding to a sequence included in the cloning vector. The resulting double stranded DNA is transformed into bacteria. cDNAs containing the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are identified by colony PCR or colony hybridization.

Once the upstream genomic sequences have been cloned and sequenced as described above, prospective promoters and transcription start sites within the upstream sequences may be identified by comparing the sequences upstream of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids with databases containing known transcription start sites, transcription factor binding sites, or promoter sequences.

In addition, promoters in the upstream sequences may be identified using promoter reporter vectors as described in Example 54.

#### EXAMPLE 54

##### Identification of Promoters in Cloned Upstream Sequences

The genomic sequences upstream of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are cloned into a suitable promoter reporter vector, such as the pSEAP-Basic, pSEAP-Enhancer, p $\beta$ -gal-Basic, p $\beta$ -gal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech. Briefly, each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase,  $\beta$ -galactosidase, or green fluorescent

protein. The sequences upstream of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be cloned into vectors which contain an enhancer for augmenting transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert indicates that a promoter sequence is present in the inserted upstream sequence.

Appropriate host cells for the promoter reporter vectors may be chosen based on the results of the above described determination of expression patterns of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. For example, if the expression pattern analysis indicates that the mRNA corresponding to a particular EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids is expressed in fibroblasts, the promoter reporter vector may be introduced into a human fibroblast cell line.

Promoter sequences within the upstream genomic DNA may be further defined by constructing nested deletions in the upstream DNA using conventional techniques such as Exonuclease III digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has reduced or obliterated promoter activity. In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into the cloning sites in the promoter reporter vectors.

### EXAMPLE 55

#### Cloning and Identification of Promoters

Using the method described in Example 54 above with 5' ESTs, sequences upstream of several genes were obtained. Using the primer pairs GGG AAG ATG GAG ATA GTA TTG CCT G (SEQ ID NO:15) and CTG CCA TGT ACA TGA TAG AGA GAT TC (SEQ ID NO:16), the promoter having the internal designation P13H2 (SEQ ID NO:17) was obtained.

Using the primer pairs GTA CCA GGGG ACT GTG ACC ATT GC (SEQ ID NO:18) and CTG TGA CCA TTG CTC CCA AGA GAG (SEQ ID NO:19), the promoter having the internal designation P15B4 (SEQ ID NO:20) was obtained.

Using the primer pairs CTG GGA TGG AAG GCA CGG TA (SEQ ID NO:21) and GAG ACC ACA CAG CTA GAC AA (SEQ ID NO:22), the promoter having the internal designation P29B6 (SEQ ID NO:23) was obtained.

Figure 4 provides a schematic description of the promoters isolated and the way they are assembled with the corresponding 5' tags. The upstream sequences were screened for the presence of motifs resembling transcription factor binding sites or known transcription start sites using the computer program MatInspector release 2.0, August 1996.

Figure 5 describes the transcription factor binding sites present in each of these promoters. The columns labeled matrix provides the name of the MatInspector matrix used. The column labeled position provides the 5' position of the promoter site. Numeration of the sequence starts from the transcription site as determined by matching the genomic sequence with the 5' EST sequence. The column labeled "orientation" indicates the DNA strand on which the site is found, with the + strand being the coding strand as determined by matching the genomic sequence with the sequence of the 5' EST. The column labeled "score" provides the MatInspector score found for this site. The column labeled "length" provides the length of the site in nucleotides. The column labeled "sequence" provides the sequence of the site found.

Bacterial clones containing plasmids containing the promoter sequences described above described above are presently stored in the inventor's laboratories under the internal identification numbers provided above. The inserts may be recovered from the deposited materials by growing an aliquot of the appropriate bacterial clone in the appropriate medium. The plasmid DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the inserted EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. The PCR product which corresponds to the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids can then be manipulated using standard cloning techniques familiar to those skilled in the art.

The promoters and other regulatory sequences located upstream of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be used to design expression vectors capable of directing the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative manner. A promoter capable of directing the desired spatial, temporal, developmental, and quantitative patterns may be selected using the results of the expression analysis described above. For example, if a promoter which confers a high level of expression in muscle is desired, the promoter sequence upstream of EST-related nucleic acids,

positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids derived from an mRNA which are expressed at a high level in muscle, as determined by the methods above, may be used in the expression vector.

Preferably, the desired promoter is placed near multiple restriction sites to facilitate the cloning of the desired insert downstream of the promoter, such that the promoter is able to drive expression of the inserted gene. The promoter may be inserted in conventional nucleic acid backbones designed for extrachromosomal replication, integration into the host chromosomes or transient expression. Suitable backbones for the present expression vectors include retroviral backbones, backbones from eukaryotic episomes such as SV40 or Bovine Papilloma Virus, backbones from bacterial episomes, or artificial chromosomes.

Preferably, the expression vectors also include a polyA signal downstream of the multiple restriction sites for directing the polyadenylation of mRNA transcribed from the gene inserted into the expression vector.

Following the identification of promoter sequences, proteins which interact with the promoter may be identified as described in Example 56 below.

#### EXAMPLE 56

##### Identification of Proteins Which Interact with Promoter Sequences, Upstream Regulatory Sequences, or mRNA

Sequences within the promoter region which are likely to bind transcription factors may be identified by homology to known transcription factor binding sites or through conventional mutagenesis or deletion analyses of reporter plasmids containing the promoter sequence. For example, deletions may be made in a reporter plasmid containing the promoter sequence of interest operably linked to an assayable reporter gene. The reporter plasmids carrying various deletions within the promoter region are transfected into an appropriate host cell and the effects of the deletions on expression levels is assessed. Transcription factor binding sites within the regions in which deletions reduce expression levels may be further localized using site directed mutagenesis, linker scanning analysis, or other techniques familiar to those skilled in the art.

Nucleic acids encoding proteins which interact with sequences in the promoter may be identified using one-hybrid systems such as those described in the manual accompanying the Matchmaker One-Hybrid System kit available from Clontech (Catalog No. K1603-1). Briefly, the Matchmaker One-hybrid system is used as follows. The target sequence for which it is desired to identify binding proteins is cloned upstream of a selectable reporter gene and integrated into the yeast genome. Preferably, multiple copies of the target sequences are inserted into the reporter plasmid in tandem. A library comprised of fusions between cDNAs to be evaluated for the ability to bind to the promoter and the activation domain of a yeast transcription factor, such as GAL4, is transformed into the yeast strain containing the integrated reporter sequence. The yeast are plated on selective media to select cells

expressing the selectable marker linked to the promoter sequence. The colonies which grow on the selective media contain genes encoding proteins which bind the target sequence. The inserts in the genes encoding the fusion proteins are further characterized by sequencing. In addition, the inserts may be inserted into expression vectors or *in vitro* transcription vectors. Binding of the polypeptides encoded by the inserts to the promoter DNA may be confirmed by techniques familiar to those skilled in the art, such as gel shift analysis or DNase protection analysis.

### **VIII. Use of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids in Gene Therapy**

The present invention also comprises the use of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids in gene therapy strategies, including antisense and triple helix strategies as described in Examples 57 and 58 below. In antisense approaches, nucleic acid sequences complementary to an mRNA are hybridized to the mRNA intracellularly, thereby blocking the expression of the protein encoded by the mRNA. The antisense sequences may prevent gene expression through a variety of mechanisms. For example, the antisense sequences may inhibit the ability of ribosomes to translate the mRNA. Alternatively, the antisense sequences may block transport of the mRNA from the nucleus to the cytoplasm, thereby limiting the amount of mRNA available for translation. Another mechanism through which antisense sequences may inhibit gene expression is by interfering with mRNA splicing. In yet another strategy, the antisense nucleic acid may be incorporated in a ribozyme capable of specifically cleaving the target mRNA.

#### **EXAMPLE 57**

##### Preparation and Use of Antisense Oligonucleotides

The antisense nucleic acid molecules to be used in gene therapy may be either DNA or RNA sequences. They may comprise a sequence complementary to the sequence of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex with sufficient stability to inhibit the expression of the mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green *et al.*, *Ann. Rev. Biochem.* 55:569-597 (1986) and Izant and Weintraub, *Cell* 36:1007-1015 (1984).

In some strategies, antisense molecules are obtained from a nucleotide sequence encoding a protein by reversing the orientation of the coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using *in vitro* transcription systems such as those which employ T7 or SP6 polymerase to

generate the transcript. Another approach involves transcription of the antisense nucleic acids *in vivo* by operably linking DNA containing the antisense sequence to a promoter in an expression vector.

Alternatively, oligonucleotides which are complementary to the strand normally transcribed in the cell may be synthesized *in vitro*. Thus, the antisense nucleic acids are complementary to the  
5 corresponding mRNA and are capable of hybridizing to the mRNA to create a duplex. In some embodiments, the antisense sequences may contain modified sugar phosphate backbones to increase stability and make them less sensitive to RNase activity. Examples of modifications suitable for use in antisense strategies are described by Rossi *et al.*, *Pharmacol. Ther.* 50(2):245-254, (1991).

Various types of antisense oligonucleotides complementary to the sequence of the EST-related  
10 nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be used. In one preferred embodiment, stable and semi-stable antisense oligonucleotides described in International Application No. PCT WO94/23026 are used. In these molecules, the 3' end or both the 3' and 5' ends are engaged in intramolecular hydrogen bonding between complementary base pairs. These molecules are better able to withstand exonuclease attacks  
15 and exhibit increased stability compared to conventional antisense oligonucleotides.

In another preferred embodiment, the antisense oligodeoxynucleotides against herpes simplex virus types 1 and 2 described in International Application No. WO 95/04141 are used.

In yet another preferred embodiment, the covalently cross-linked antisense oligonucleotides described in International Application No. WO 96/31523 are used. These double- or single-stranded  
20 oligonucleotides comprise one or more, respectively, inter- or intra-oligonucleotide covalent cross-linkages, wherein the linkage consists of an amide bond between a primary amine group of one strand and a carboxyl group of the other strand or of the same strand, respectively, the primary amine group being directly substituted in the 2' position of the strand nucleotide monosaccharide ring, and the carboxyl group being carried by an aliphatic spacer group substituted on a nucleotide or nucleotide  
25 analog of the other strand or the same strand, respectively.

The antisense oligodeoxynucleotides and oligonucleotides disclosed in International Application No. WO 92/18522 may also be used. These molecules are stable to degradation and contain at least one transcription control recognition sequence which binds to control proteins and are effective as decoys therefor. These molecules may contain "hairpin" structures, "dumbbell" structures, "modified  
30 dumbbell" structures, "cross-linked" decoy structures and "loop" structures.

In another preferred embodiment, the cyclic double-stranded oligonucleotides described in European Patent Application No. 0 572 287 A2. These ligated oligonucleotide "dumbbells" contain the binding site for a transcription factor and inhibit expression of the gene under control of the transcription factor by sequestering the factor.

35 Use of the closed antisense oligonucleotides disclosed in International Application No. WO 92/19732 is also contemplated. Because these molecules have no free ends, they are more resistant to

degradation by exonucleases than are conventional oligonucleotides. These oligonucleotides may be multifunctional, interacting with several regions which are not adjacent to the target mRNA.

The appropriate level of antisense nucleic acids required to inhibit gene expression may be determined using *in vitro* expression analysis. The antisense molecule may be introduced into the cells  
5 by diffusion, injection, infection or transfection using procedures known in the art. For example, the antisense nucleic acids can be introduced into the body as a bare or naked oligonucleotide, oligonucleotide encapsulated in lipid, oligonucleotide sequence encapsidated by viral protein, or as an oligonucleotide operably linked to a promoter contained in an expression vector. The expression vector may be any of a variety of expression vectors known in the art, including retroviral or viral vectors,  
10 vectors capable of extrachromosomal replication, or integrating vectors. The vectors may be DNA or RNA.

The antisense molecules are introduced onto cell samples at a number of different concentrations preferably between  $1 \times 10^{-10}$ M to  $1 \times 10^{-4}$ M. Once the minimum concentration that can adequately control gene expression is identified, the optimized dose is translated into a dosage suitable  
15 for use *in vivo*. For example, an inhibiting concentration in culture of  $1 \times 10^{-7}$  translates into a dose of approximately 0.6 mg/kg bodyweight. Levels of oligonucleotide approaching 100 mg/kg bodyweight or higher may be possible after testing the toxicity of the oligonucleotide in laboratory animals. It is additionally contemplated that cells from the vertebrate are removed, treated with the antisense oligonucleotide, and reintroduced into the vertebrate.

20 It is further contemplated that the antisense oligonucleotide sequence is incorporated into a ribozyme sequence to enable the antisense to specifically bind and cleave its target mRNA. For technical applications of ribozyme and antisense oligonucleotides see Rossi *et al.*, *supra*.

In a preferred application of this invention, the polypeptide encoded by the gene is first identified, so that the effectiveness of antisense inhibition on translation can be monitored using  
25 techniques that include but are not limited to antibody-mediated tests such as RIAs and ELISA, functional assays, or radiolabeling.

The EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may also be used in gene therapy approaches based on intracellular triple helix formation. Triple helix oligonucleotides are used to inhibit transcription from a  
30 genome. They are particularly useful for studying alterations in cell activity as it is associated with a particular gene. The EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids of the present invention or, more preferably, a portion of those sequences, can be used to inhibit gene expression in individuals having diseases associated with expression of a particular gene. Similarly, the EST-related nucleic acids,  
35 positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids can be used to study the effect of inhibiting transcription of a particular gene within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies.

However, homopyrimidine sequences can also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. Thus, both types of sequences from the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are contemplated within the scope of this invention.

### EXAMPLE 58

#### Preparation and use of Triple Helix Probes

The sequences of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting gene expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting gene expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which normally express the target gene. The oligonucleotides may be prepared on an oligonucleotide synthesizer or they may be purchased commercially from a company specializing in custom oligonucleotide synthesis, such as GENSET, Paris, France.

The oligonucleotides may be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced gene expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the target gene in cells which have been treated with the oligonucleotide. The cell functions to be monitored are predicted based upon the homologies of the target genes corresponding to the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids from which the oligonucleotide were derived with known gene sequences that have been associated with a particular function. The cell functions can also be predicted based on the presence of abnormal physiologies within cells derived from individuals with a particular inherited disease, particularly when the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are associated with the disease using techniques described herein.

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced *in vivo* using the techniques described above and in Example 56 at a dosage calculated based on the *in vitro* results, as described in Example 57.

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to



stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin *et al.* (*Science* 245:967-971 (1989)).

### EXAMPLE 59

5                    Use of EST-related nucleic acids, positional segments of  
                     EST-related nucleic acids or fragments of positional segments of  
                     EST-related nucleic acids to express an Encoded Protein in a Host Organism

The EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may also be used to express an encoded protein or  
10 polypeptide in a host organism to produce a beneficial effect. In addition, nucleic acids encoding the EST-related polypeptides, positional segments of EST-related polypeptides or fragments of positional segments of EST-related polypeptides may be used to express the encoded protein or polypeptide in a host organism to produce a beneficial effect.

In such procedures, the encoded protein or polypeptide may be transiently expressed in the host  
15 organism or stably expressed in the host organism. The encoded protein or polypeptide may have any of the activities described above. The encoded protein or polypeptide may be a protein or polypeptide which the host organism lacks or, alternatively, the encoded protein may augment the existing levels of the protein in the host organism.

In some embodiments in which the protein or polypeptide is secreted, nucleic acids encoding the  
20 full length protein (*i.e.* the signal peptide and the mature protein), or nucleic acids encoding only the mature protein (*i.e.* the protein generated when the signal peptide is cleaved off) is introduced into the host organism.

The nucleic acids encoding the proteins or polypeptides may be introduced into the host organism using a variety of techniques known to those of skill in the art. For example, the extended  
25 cDNA may be injected into the host organism as naked DNA such that the encoded protein is expressed in the host organism, thereby producing a beneficial effect.

Alternatively, the nucleic acids encoding the protein or polypeptide may be cloned into an expression vector downstream of a promoter which is active in the host organism. The expression vector may be any of the expression vectors designed for use in gene therapy, including viral or retroviral  
30 vectors. The expression vector may be directly introduced into the host organism such that the encoded protein is expressed in the host organism to produce a beneficial effect. In another approach, the expression vector may be introduced into cells *in vitro*. Cells containing the expression vector are thereafter selected and introduced into the host organism, where they express the encoded protein or polypeptide to produce a beneficial effect.

35

### EXAMPLE 60

Use of Signal Peptides To Import Proteins Into Cells

The short core hydrophobic region (h) of signal peptides encoded by the sequences of SEQ ID NOs. 24-728 and 766-792 may also be used as a carrier to import a peptide or a protein of interest, so-called cargo, into tissue culture cells (Lin *et al.*, *J. Biol. Chem.*, 270: 14225-14258 (1995); Du *et al.*, *J. Peptide Res.*, 51: 235-243 (1998); Rojas *et al.*, *Nature Biotech.*, 16: 370-375 (1998)).

5 When cell permeable peptides of limited size (approximately up to 25 amino acids) are to be translocated across cell membrane, chemical synthesis may be used in order to add the h region to either the C-terminus or the N-terminus to the cargo peptide of interest. Alternatively, when longer peptides or proteins are to be imported into cells, nucleic acids can be genetically engineered, using techniques familiar to those skilled in the art, in order to link the extended cDNA sequence encoding the h region to  
10 the 5' or the 3' end of a DNA sequence coding for a cargo polypeptide. Such genetically engineered nucleic acids are then translated either *in vitro* or *in vivo* after transfection into appropriate cells, using conventional techniques to produce the resulting cell permeable polypeptide. Suitable hosts cells are then simply incubated with the cell permeable polypeptide which is then translocated across the membrane.

15 This method may be applied to study diverse intracellular functions and cellular processes. For instance, it has been used to probe functionally relevant domains of intracellular proteins and to examine protein-protein interactions involved in signal transduction pathways (Lin *et al.*, *supra*; Lin *et al.*, *J. Biol. Chem.*, 271: 5305-5308 (1996); Rojas *et al.*, *J. Biol. Chem.*, 271: 27456-27461 (1996); Liu *et al.*, *Proc. Natl. Acad. Sci. USA*, 93: 11819-11824 (1996); Rojas *et al.*, *Bioch. Biophys. Res. Commun.*, 234: 675-  
20 680 (1997)).

Such techniques may be used in cellular therapy to import proteins producing therapeutic effects. For instance, cells isolated from a patient may be treated with imported therapeutic proteins and then re-introduced into the host organism.

Alternatively, the h region of signal peptides of the present invention could be used in  
25 combination with a nuclear localization signal to deliver nucleic acids into cell nucleus. Such oligonucleotides may be antisense oligonucleotides or oligonucleotides designed to form triple helixes, as described above, in order to inhibit processing and maturation of a target cellular RNA.

### EXAMPLE 61

#### 30 Computer Embodiments

As used herein the term "nucleic acid codes of SEQ ID NOs. 24-811 and 1600-1622" encompasses the nucleotide sequences of SEQ ID NOs. 24-811 and 1600-1622, fragments of SEQ ID NOs. 24-811 and 1600-1622, nucleotide sequences homologous to SEQ ID NOs. 24-811 and 1600-1622 or homologous to fragments of SEQ ID NOs. 24-811 and 1600-1622, and sequences  
35 complementary to all of the preceding sequences. The fragments include portions of SEQ ID NOs. 24-811 and 1600-1622 comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive nucleotides of SEQ ID NOs. 24-811 and 1600-1622. Preferably, the fragments are novel

fragments. Preferably the fragments include polynucleotides described in Table II, polynucleotides described in Table III, polynucleotides described in Table IV or portions thereof comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive nucleotides of the polynucleotides described in Tables II, III, or IV. Homologous sequences and fragments of SEQ ID

5 NOs. 24-811 and 1600-1622 refer to a sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, or 75% homology to these sequences. Homology may be determined using any of the computer programs and parameters described in Example 18, including BLAST2N with the default parameters or with any modified parameters. Homologous sequences also include RNA sequences in which uridines replace the thymines in the nucleic acid codes of SEQ ID NOs. 24-811 and 1600-1622. The

10 homologous sequences may be obtained using any of the procedures described herein or may result from the correction of a sequencing error as described above. Preferably the homologous sequences and fragments of SEQ ID NOs. 24-811 and 1600-1622 include polynucleotides described in Table II, polynucleotides described in Table III, polynucleotides described in Table IV or portions thereof comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive

15 nucleotides of the polynucleotides described in Tables II, III, or IV. It will be appreciated that the nucleic acid codes of SEQ ID NOs. 24-811 and 1600-1622 can be represented in the traditional single character format (See the inside back cover of Styer, Lubert. *Biochemistry*, 3<sup>rd</sup> edition. W. H Freeman & Co., New York.) or in any other format which records the identity of the nucleotides in a sequence.

As used herein the term "polypeptide codes of SEQ ID NOS. 812-1599" encompasses the

20 polypeptide sequence of SEQ ID NOs. 812-1599 which are encoded by the 5' EST s of SEQ ID NOs. 24-811 and 1600-1622, polypeptide sequences homologous to the polypeptides of SEQ ID NOS. 812-1599, or fragments of any of the preceding sequences. Homologous polypeptide sequences refer to a polypeptide sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75% homology to one of the polypeptide sequences of SEQ ID NOS. 812-1599. Homology may be determined using any

25 of the computer programs and parameters described herein, including FASTA with the default parameters or with any modified parameters. The homologous sequences may be obtained using any of the procedures described herein or may result from the correction of a sequencing error as described above. The polypeptide fragments comprise at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids of the polypeptides of SEQ ID NOS. 812-1599. Preferably, the fragments are

30 novel fragments. Preferably, the fragments include polypeptides encoded by the polynucleotides described in Table II, or portions thereof comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids of the polypeptides encoded by the polynucleotides described in Table II. It will be appreciated that the polypeptide codes of the SEQ ID NOS. 812-1599 can be represented in the traditional single character format or three letter format (See the inside back cover of Starrier, Lubert.

35 *Biochemistry*, 3<sup>rd</sup> edition. W. H Freeman & Co., New York.) or in any other format which relates the identity of the polypeptides in a sequence.

It will be appreciated by those skilled in the art that the nucleic acid codes of SEQ ID NOs. 24-811 and 1600-1622 and polypeptide codes of SEQ ID NOS. 812-1599 can be stored, recorded, and manipulated on any medium which can be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. A skilled  
5 artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid codes of SEQ ID NOs. 24-811 and 1600-1622, one or more of the polypeptide codes of SEQ ID NOS. 812-1599. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 nucleic acid codes of SEQ ID NOs. 24-811 and 1600-1622. Another  
10 aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 polypeptide codes of SEQ ID NOS. 812-1599.

Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disk, a floppy disk, a magnetic tape, CD-ROM, Digital Versatile Disk (DVD), Random  
15 Access Memory (RAM), or Read Only Memory (ROM) as well as other types of other media known to those skilled in the art.

Embodiments of the present invention include systems, particularly computer systems which store and manipulate the sequence information described herein. One example of a computer system 100 is illustrated in block diagram form in Figure 6. As used herein, "a computer system" refers to the  
20 hardware components, software components, and data storage components used to analyze the nucleotide sequences of the nucleic acid codes of SEQ ID NOs. 24-811 and 1600-1622, or the amino acid sequences of the polypeptide codes of SEQ ID NOS. 812-1599. In one embodiment, the computer system 100 is a Sun Enterprise 1000 server (Sun Microsystems, Palo Alto, CA). The computer system 100 preferably includes a processor for processing, accessing and manipulating the sequence data. The  
25 processor 105 can be any well-known type of central processing unit, such as the Pentium III from Intel Corporation, or similar processor from Sun, Motorola, Compaq or International Business Machines.

Preferably, the computer system 100 is a general purpose system that comprises the processor 105 and one or more internal data storage components 110 for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can  
30 readily appreciate that any one of the currently available computer systems are suitable.

In one particular embodiment, the computer system 100 includes a processor 105 connected to a bus which is connected to a main memory 115 (preferably implemented as RAM) and one or more internal data storage devices 110, such as a hard drive and/or other computer readable media having data recorded thereon. In some embodiments, the computer system 100 further includes one or more data  
35 retrieving device 118 for reading the data stored on the internal data storage devices 110.

The data retrieving device 118 may represent, for example, a floppy disk drive, a compact disk drive, a magnetic tape drive, etc. In some embodiments, the internal data storage device 110 is a

removable computer readable medium such as a floppy disk, a compact disk, a magnetic tape, etc. containing control logic and/or data recorded thereon. The computer system 100 may advantageously include or be programmed by appropriate software for reading the control logic and/or the data from the data storage component once inserted in the data retrieving device.

5       The computer system 100 includes a display 120 which is used to display output to a computer user. It should also be noted that the computer system 100 can be linked to other computer systems 125a-c in a network or wide area network to provide centralized access to the computer system 100.

Software for accessing and processing the nucleotide sequences of the nucleic acid codes of SEQ ID NOS. 24-811 and 1600-1622, or the amino acid sequences of the polypeptide codes of SEQ ID  
10 NOS. 812-1599 (such as search tools, compare tools, and modeling tools etc.) may reside in main memory 115 during execution.

In some embodiments, the computer system 100 may further comprise a sequence comparer for comparing the above-described nucleic acid codes of SEQ ID NOS. 24-811 and 1600-1622 or polypeptide codes of SEQ ID NOS. 812-1599 stored on a computer readable medium to reference  
15 nucleotide or polypeptide sequences stored on a computer readable medium. A "sequence comparer" refers to one or more programs which are implemented on the computer system 100 to compare a nucleotide or polypeptide sequence with other nucleotide or polypeptide sequences and/or compounds including but not limited to peptides, peptidomimetics, and chemicals stored within the data storage means. For example, the sequence comparer may compare the nucleotide sequences of the nucleic acid  
20 codes of SEQ ID NOS. 24-811 and 1600-1622, or the amino acid sequences of the polypeptide codes of SEQ ID NOS. 812-1599 stored on a computer readable medium to reference sequences stored on a computer readable medium to identify homologies, motifs implicated in biological function, or structural motifs. The various sequence comparer programs identified elsewhere in this patent specification are particularly contemplated for use in this aspect of the invention.

25       Figure 7 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database. The database of sequences can be a private database stored within the computer system 100, or a public database such as GENBANK, PIR OR SWISSPROT that is available through the Internet.

30       The process 200 begins at a start state 201 and then moves to a state 202 wherein the new sequence to be compared is stored to a memory in a computer system 100. As discussed above, the memory could be any type of memory, including RAM or an internal storage device.

The process 200 then moves to a state 204 wherein a database of sequences is opened for analysis and comparison. The process 200 then moves to a state 206 wherein the first sequence stored in  
35 the database is read into a memory on the computer. A comparison is then performed at a state 210 to determine if the first sequence is the same as the second sequence. It is important to note that this step is not limited to performing an exact comparison between the new sequence and the first sequence in the

database. Well-known methods are known to those of skill in the art for comparing two nucleotide or protein sequences, even if they are not identical. For example, gaps can be introduced into one sequence in order to raise the homology level between the two tested sequences. The parameters that control whether gaps or other features are introduced into a sequence during comparison are normally entered by the user of the computer system.

Once a comparison of the two sequences has been performed at the state 210, a determination is made at a decision state 210 whether the two sequences are the same. Of course, the term "same" is not limited to sequences that are absolutely identical. Sequences that are within the homology parameters entered by the user will be marked as "same" in the process 200.

If a determination is made that the two sequences are the same, the process 200 moves to a state 214 wherein the name of the sequence from the database is displayed to the user. This state notifies the user that the sequence with the displayed name fulfills the homology constraints that were entered. Once the name of the stored sequence is displayed to the user, the process 200 moves to a decision state 218 wherein a determination is made whether more sequences exist in the database. If no more sequences exist in the database, then the process 200 terminates at an end state 220. However, if more sequences do exist in the database, then the process 200 moves to a state 224 wherein a pointer is moved to the next sequence in the database so that it can be compared to the new sequence. In this manner, the new sequence is aligned and compared with every sequence in the database.

It should be noted that if a determination had been made at the decision state 212 that the sequences were not homologous, then the process 200 would move immediately to the decision state 218 in order to determine if any other sequences were available in the database for comparison.

Accordingly, one aspect of the present invention is a computer system comprising a processor, a data storage device having stored thereon a nucleic acid code of SEQ ID NOs. 24-811 and 1600-1622 or a polypeptide code of SEQ ID NOS. 812-1599, a data storage device having retrievably stored thereon reference nucleotide sequences or polypeptide sequences to be compared to the nucleic acid code of SEQ ID NOs. 24-811 and 1600-1622 or polypeptide code of SEQ ID NOS. 812-1599 and a sequence comparer for conducting the comparison. The sequence comparer may indicate a homology level between the sequences compared or identify structural motifs in the above described nucleic acid code of SEQ ID NOs. 24-811 and 1600-1622 and polypeptide codes of SEQ ID NOS. 812-1599 or it may identify structural motifs in sequences which are compared to these nucleic acid codes and polypeptide codes. In some embodiments, the data storage device may have stored thereon the sequences of at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of SEQ ID NOs. 24-811 and 1600-1622 or polypeptide codes of SEQ ID NOS. 812-1599.

Another aspect of the present invention is a method for determining the level of homology between a nucleic acid code of SEQ ID NOs. 24-811 and 1600-1622 and a reference nucleotide sequence, comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through the use of a computer program which determines homology levels and determining homology

between the nucleic acid code and the reference nucleotide sequence with the computer program. The computer program may be any of a number of computer programs for determining homology levels, including those specifically enumerated herein, including BLAST2N with the default parameters or with any modified parameters. The method may be implemented using the computer systems described above. The method may also be performed by reading 2, 5, 10, 15, 20, 25, 30, or 50 of the above described nucleic acid codes of SEQ ID NOs. 24-811 and 1600-1622 through use of the computer program and determining homology between the nucleic acid codes and reference nucleotide sequences.

Figure 8 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous. The process 250 begins at a start state 252 and then moves to a state 254 wherein a first sequence to be compared is stored to a memory. The second sequence to be compared is then stored to a memory at a state 256. The process 250 then moves to a state 260 wherein the first character in the first sequence is read and then to a state 262 wherein the first character of the second sequence is read. It should be understood that if the sequence is a nucleotide sequence, then the character would normally be either A, T, C, G or U. If the sequence is a protein sequence, then it should be in the single letter amino acid code so that the first and sequence sequences can be easily compared.

A determination is then made at a decision state 264 whether the two characters are the same. If they are the same, then the process 250 moves to a state 268 wherein the next characters in the first and second sequences are read. A determination is then made whether the next characters are the same. If they are, then the process 250 continues this loop until two characters are not the same. If a determination is made that the next two characters are not the same, the process 250 moves to a decision state 274 to determine whether there are any more characters either sequence to read.

If there aren't any more characters to read, then the process 250 moves to a state 276 wherein the level of homology between the first and second sequences is displayed to the user. The level of homology is determined by calculating the proportion of characters between the sequences that were the same out of the total number of sequences in the first sequence. Thus, if every character in a first 100 nucleotide sequence aligned with a every character in a second sequence, the homology level would be 100%.

Alternatively, the computer program may be a computer program which compares the nucleotide sequences of the nucleic acid codes of the present invention, to reference nucleotide sequences in order to determine whether the nucleic acid code of SEQ ID NOs. 24-811 and 1600-1622 differs from a reference nucleic acid sequence at one or more positions. Optionally such a program records the length and identity of inserted, deleted or substituted nucleotides with respect to the sequence of either the reference polynucleotide or the nucleic acid code of SEQ ID NOs. 24-811 and 1600-1622. In one embodiment, the computer program may be a program which determines whether the nucleotide sequences of the nucleic acid codes of SEQ ID NOs. 24-811 and 1600-1622 contain a biallelic marker

or single nucleotide polymorphism (SNP) with respect to a reference nucleotide sequence. This single nucleotide polymorphism may comprise a single base substitution, insertion, or deletion, while this biallelic marker may comprise about one to ten consecutive bases substituted, inserted or deleted.

Another aspect of the present invention is a method for determining the level of homology  
5 between a polypeptide code of SEQ ID NOS. 812-1599 and a reference polypeptide sequence, comprising the steps of reading the polypeptide code of SEQ ID NOS. 812-1599 and the reference polypeptide sequence through use of a computer program which determines homology levels and determining homology between the polypeptide code and the reference polypeptide sequence using the computer program.

10 Accordingly, another aspect of the present invention is a method for determining whether a nucleic acid code of SEQ ID NOS. 24-811 and 1600-1622 differs at one or more nucleotides from a reference nucleotide sequence comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through use of a computer program which identifies differences between nucleic acid sequences and identifying differences between the nucleic acid code and the reference nucleotide  
15 sequence with the computer program. In some embodiments, the computer program is a program which identifies single nucleotide polymorphisms. The method may be implemented by the computer systems described above and the method illustrated in Figure 8. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of SEQ ID NOS. 24-811 and 1600-1622 and the reference nucleotide sequences through the use of the computer program and identifying differences  
20 between the nucleic acid codes and the reference nucleotide sequences with the computer program.

In other embodiments the computer based system may further comprise an identifier for identifying features within the nucleotide sequences of the nucleic acid codes of SEQ ID NOS. 24-811 and 1600-1622 or the amino acid sequences of the polypeptide codes of SEQ ID NOS. 812-1599.

An "identifier" refers to one or more programs which identifies certain features within the  
25 above-described nucleotide sequences of the nucleic acid codes of SEQ ID NOS. 24-811 and 1600-1622 or the amino acid sequences of the polypeptide codes of SEQ ID NOS. 812-1599. In one embodiment, the identifier may comprise a program which identifies an open reading frame in the cDNAs codes of SEQ ID NOS. 24-811 and 1600-1622.

Figure 9 is a flow diagram illustrating one embodiment of an identifier process 300 for  
30 detecting the presence of a feature in a sequence. The process 300 begins at a start state 302 and then moves to a state 304 wherein a first sequence that is to be checked for features is stored to a memory 115 in the computer system 100. The process 300 then moves to a state 306 wherein a database of sequence features is opened. Such a database would include a list of each feature's attributes along with the name of the feature. For example, a feature name could be "Initiation Codon" and the  
35 attribute would be "ATG". Another example would be the feature name "TAATAA Box" and the feature attribute would be "TAATAA". An example of such a database is produced by the University of Wisconsin Genetics Computer Group ([www.gcg.com](http://www.gcg.com)).



Once the database of features is opened at the state 306, the process 300 moves to a state 308 wherein the first feature is read from the database. A comparison of the attribute of the first feature with the first sequence is then made at a state 310. A determination is then made at a decision state 316 whether the attribute of the feature was found in the first sequence. If the attribute was found, then the process 300 moves to a state 318 wherein the name of the found feature is displayed to the user.

The process 300 then moves to a decision state 320 wherein a determination is made whether move features exist in the database. If no more features do exist, then the process 300 terminates at an end state 324. However, if more features do exist in the database, then the process 300 reads the next sequence feature at a state 326 and loops back to the state 310 wherein the attribute of the next feature is compared against the first sequence.

It should be noted, that if the feature attribute is not found in the first sequence at the decision state 316, the process 300 moves directly to the decision state 320 in order to determine if any more features exist in the database.

In another embodiment, the identifier may comprise a molecular modeling program which determines the 3-dimensional structure of the polypeptides codes of SEQ ID NOS. 812-1599. In some embodiments, the molecular modeling program identifies target sequences that are most compatible with profiles representing the structural environments of the residues in known three-dimensional protein structures. (See, e.g., Eisenberg et al., U.S. Patent No. 5,436,850 issued July 25, 1995). In another technique, the known three-dimensional structures of proteins in a given family are superimposed to define the structurally conserved regions in that family. This protein modeling technique also uses the known three-dimensional structure of a homologous protein to approximate the structure of the polypeptide codes of SEQ ID NOS. 812-1599. (See e.g., Srinivasan, et al., U.S. Patent No. 5,557,535 issued September 17, 1996). Conventional homology modeling techniques have been used routinely to build models of proteases and antibodies. (Sowdhamini et al., Protein Engineering 10:207, 215 (1997)). Comparative approaches can also be used to develop three-dimensional protein models when the protein of interest has poor sequence identity to template proteins. In some cases, proteins fold into similar three-dimensional structures despite having very weak sequence identities. For example, the three-dimensional structures of a number of helical cytokines fold in similar three-dimensional topology in spite of weak sequence homology.

The recent development of threading methods now enables the identification of likely folding patterns in a number of situations where the structural relatedness between target and template(s) is not detectable at the sequence level. Hybrid methods, in which fold recognition is performed using Multiple Sequence Threading (MST), structural equivalencies are deduced from the threading output using a distance geometry program DRAGON to construct a low resolution model, and a full-atom representation is constructed using a molecular modeling package such as QUANTA.

According to this 3-step approach, candidate templates are first identified by using the novel fold recognition algorithm MST, which is capable of performing simultaneous threading of multiple aligned sequences onto one or more 3-D structures. In a second step, the structural equivalencies obtained from the MST output are converted into interresidue distance restraints and fed into the distance geometry program DRAGON, together with auxiliary information obtained from secondary structure predictions. The program combines the restraints in an unbiased manner and rapidly generates a large number of low resolution model confirmations. In a third step, these low resolution model confirmations are converted into full-atom models and subjected to energy minimization using the molecular modeling package QUANTA. (See e.g., Aszódi et al., *Proteins: Structure, Function, and Genetics*, Supplement 1:38-42 (1997)).

The results of the molecular modeling analysis may then be used in rational drug design techniques to identify agents which modulate the activity of the polypeptide codes of SEQ ID NOS. 812-1599.

Accordingly, another aspect of the present invention is a method of identifying a feature within the nucleic acid codes of SEQ ID NOS. 24-811 and 1600-1622 or the polypeptide codes of SEQ ID NOS. 812-1599 comprising reading the nucleic acid code(s) or the polypeptide code(s) through the use of a computer program which identifies features therein and identifying features within the nucleic acid code(s) or polypeptide code(s) with the computer program. In one embodiment, computer program comprises a computer program which identifies open reading frames. In a further embodiment, the computer program identifies structural motifs in a polypeptide sequence. In another embodiment, the computer program comprises a molecular modeling program. The method may be performed by reading a single sequence or at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of SEQ ID NOS. 24-811 and 1600-1622 or the polypeptide codes of SEQ ID NOS. 812-1599 through the use of the computer program and identifying features within the nucleic acid codes or polypeptide codes with the computer program.

The nucleic acid codes of SEQ ID NOS. 24-811 and 1600-1622 or the polypeptide codes of SEQ ID NOS. 812-1599 may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, the nucleic acid codes of SEQ ID NOS. 24-811 and 1600-1622 or the polypeptide codes of SEQ ID NOS. 812-1599 may be stored as text in a word processing file, such as MicrosoftWORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE. In addition, many computer programs and databases may be used as sequence comparers, identifiers, or sources of reference nucleotide or polypeptide sequences to be compared to the nucleic acid codes of SEQ ID NOS. 24-811 and 1600-1622 or the polypeptide codes of SEQ ID NOS. 812-1599. The following list is intended not to limit the invention but to provide guidance to programs and databases which are useful with the nucleic acid codes of SEQ ID NOS. 24-811 and 1600-1622 or the polypeptide codes of SEQ ID NOS. 812-1599. The programs and databases which may be used include, but are not limited to: MacPattern (EMBL),

DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul et al, *J. Mol. Biol.* 215: 403 (1990)), FASTA (Pearson and Lipman, *Proc. Natl. Acad. Sci. USA*, 85: 2444 (1988)), FASTDB (Brutlag et al. *Comp. App. Biosci.* 5 6:237-245, 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius<sup>2</sup>.DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMM (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations 10 Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), the EMBL/Swissprotein database, the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwent's World Drug Index database, the 15 BioByteMasterFile database, the Genbank database, and the Genseqn database. Many other programs and data bases would be apparent to one of skill in the art given the present disclosure.

Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, 20 sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

## EXAMPLE 62

### Methods of Making Nucleic Acids

25 The present invention also comprises methods of making the EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of the EST-related nucleic acids, or fragments of positional segments of the EST-related nucleic acids. The methods comprise sequentially linking together nucleotides to produce the nucleic acids having the preceding sequences. A variety of methods of synthesizing nucleic acids are known to those skilled in the art.

30 In many of these methods, synthesis is conducted on a solid support. These included the 3' phosphoramidite methods in which the 3' terminal base of the desired oligonucleotide is immobilized on an insoluble carrier. The nucleotide base to be added is blocked at the 5' hydroxyl and activated at the 3' hydroxyl so as to cause coupling with the immobilized nucleotide base. Deblocking of the new immobilized nucleotide compound and repetition of the cycle will produce the desired 35 polynucleotide. Alternatively, polynucleotides may be prepared as described in U.S. Patent No. 5,049,656. In some embodiments, several polynucleotides prepared as described above are ligated together to generate longer polynucleotides having a desired sequence.

**EXAMPLE 63**Methods of Making Polypeptides

The present invention also comprises methods of making the polynucleotides encoded by EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of the EST-related nucleic acids, or fragments of positional segments of the EST-related nucleic acids and methods of making the EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides, or fragments of EST-related polypeptides. The methods comprise sequentially linking together amino acids to produce the nucleic polypeptides having the preceding sequences. In some embodiments, the polypeptides made by these methods are 150 amino acid or less in length. In other embodiments, the polypeptides made by these methods are 120 amino acids or less in length.

A variety of methods of making polypeptides are known to those skilled in the art, including methods in which the carboxyl terminal amino acid is bound to polyvinyl benzene or another suitable resin. The amino acid to be added possesses blocking groups on its amino moiety and any side chain reactive groups so that only its carboxyl moiety can react. The carboxyl group is activated with carbodiimide or another activating agent and allowed to couple to the immobilized amino acid. After removal of the blocking group, the cycle is repeated to generate a polypeptide having the desired sequence. Alternatively, the methods described in U.S. Patent No. 5,049,656 may be used.

As discussed above, the EST-related nucleic acids, fragments of the EST-related nucleic acids, positional segments of the EST-related nucleic acids, or fragments of positional segments of the EST-related nucleic acids can be used for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; production of secreted polypeptides or chimeric polypeptides, antibody production, as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination for expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein or polypeptide which binds or potentially binds to another protein or polypeptide (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris *et al.*, *Cell* 75:791-803 (1993)) to identify polynucleotides encoding the other protein or polypeptide with which binding occurs or to identify inhibitors of the binding interaction.

The proteins or polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein or polypeptide binds or potentially binds to another protein or polypeptide (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other protein with which binding occurs or to identify inhibitors of the binding interaction. Proteins or polypeptides involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning; A Laboratory Manual," 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E.F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology; Guide to Molecular Cloning Techniques," Academic Press, Berger, S.L. and A.R. Kimmel eds., 1987.

Polynucleotides and proteins or polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the protein or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the protein or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

Although this invention has been described in terms of certain preferred embodiments, other embodiments which will be apparent to those of ordinary skill in the art in view of the disclosure herein are also within the scope of this invention. Accordingly, the scope of the invention is intended to be limited only by reference to the appended claims.

#### Sequence Listing Free Text

The following free text appears in the accompanying Sequence Listing:

Von Heijne matrix

score

sequence

name

martinspector prediction

CLAIMS

1. A purified nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and sequences complementary to the sequences of  
5 SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622.

2. A purified nucleic acid comprising at least 15 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and sequences complementary to the sequences of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622.  
10

3. A purified or isolated polypeptide comprising a sequence selected from the group consisting of the sequences of SEQ ID NOs. 812-1599.

4. A method of making a cDNA comprising the steps of:  
15 a) contacting a collection of mRNA molecules from human cells with a primer comprising at least 15 consecutive nucleotides of a sequence selected from the group consisting of the sequences complementary to SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622;  
b) hybridizing said primer to an mRNA in said collection that encodes said protein;  
c) reverse transcribing said hybridized primer to make a first cDNA strand from said  
20 mRNA;  
d) making a second cDNA strand complementary to said first cDNA strand; and  
e) isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

25 5. A method of making a cDNA comprising the steps of:  
a) obtaining a cDNA comprising a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622;  
b) contacting said cDNA with a detectable probe comprising at least 15 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID  
30 NOs. 1600-1622 and the sequences complementary to SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 under conditions which permit said probe to hybridize to said cDNA;  
c) identifying a cDNA which hybridizes to said detectable probe; and  
d) isolating said cDNA which hybridizes to said probe.

35 6. A method of making a cDNA comprising the steps of:  
a) contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA;  
b) hybridizing said first primer to said polyA tail;

- c) reverse transcribing said mRNA to make a first cDNA strand;
- d) making a second cDNA strand complementary to said first cDNA strand using at least one primer comprising at least 15 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622; and
- 5 e) isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

7. A method of making a polypeptide comprising the steps of:

- a) obtaining a cDNA which encodes a polypeptide encoded by a nucleic acid comprising  
10 a sequence selected from the group consisting of SEQ ID NOs. 24-811 or a cDNA which encodes a polypeptide comprising at least 10 consecutive amino acids of a polypeptide encoded by a sequence selected from the group consisting of SEQ ID NOs. 24-811;
- b) inserting said cDNA in an expression vector such that said cDNA is operably linked to a promoter;
- 15 c) introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said cDNA; and
- d) isolating said protein.

8. In an array of discrete ESTs or fragments thereof of at least 15 nucleotides in length, the  
20 improvement comprising inclusion in said array of at least one sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622, the sequences complementary to the sequences of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and fragments comprising at least 15 consecutive nucleotides of said sequence.

25 9. The array of Claim 8 including therein at least five sequences selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622, the sequences complementary to the sequences of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and fragments comprising at least 15 consecutive nucleotides of said sequences.

30 10. An enriched population of recombinant nucleic acids, said recombinant nucleic acids comprising an insert nucleic acid and a backbone nucleic acid, wherein at least 5% of said insert nucleic acids in said population comprise a sequence selected from the group consisting of SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622, the sequences complementary to SEQ ID NOs. 24-811 and SEQ ID NOs. 1600-1622 and fragments comprising at least 15 consecutive nucleotides of said  
35 sequences.

11. An antibody composition capable of selectively binding to an epitope-containing fragment of a polypeptide comprising a contiguous span of at least 8 amino acids of any of SEQ ID NOs. 812-1599, wherein said antibody is polyclonal or monoclonal.

12. A computer readable medium having stored thereon a sequence selected from the group consisting of a nucleic acid code of SEQ ID NOs. 24-811 and 1600-1622 and a polypeptide code of SEQ ID NOs. 812-1599.

5

13. A computer system comprising a processor and a data storage device wherein said data storage device has stored thereon a sequence selected from the group consisting of a nucleic acid code of SEQ ID NOs. 24-811 and 1600-1622 and a polypeptide code of SEQ ID NOs. 812-1599.

10

14. The computer system of Claim 13 further comprising a sequence comparer and a data storage device having reference sequences stored thereon.

15. The computer system of Claim 14 wherein said sequence comparer comprises a computer program which indicates polymorphisms.

15

16. The computer system of Claim 13 further comprising an identifier which identifies features in said sequence.

17. A method for comparing a first sequence to a reference sequence wherein said first sequence is selected from the group consisting of a nucleic acid code of SEQ ID NOs. 24-811 and 1600-1622 and a polypeptide code of SEQ ID NOs. 812-1599 comprising the steps of:

- a) reading said first sequence and said reference sequence through use of a computer program which compares sequences; and
- b) determining differences between said first sequence and said reference sequence with said computer program.

25

18. The method of Claim 17, wherein said step of determining differences between the first sequence and the reference sequence comprises identifying polymorphisms.

19. A method for identifying a feature in a sequence selected from the group consisting of a nucleic acid code of SEQ ID NOs. 24-811 and 1600-1622 and a polypeptide code of SEQ ID NOs. 812-1599 comprising the steps of:

- a) reading said sequence through the use of a computer program which identifies features in sequences; and
- b) identifying features in said sequence with said computer program.

35

20. A vector comprising a nucleic acid according to either Claims 1 or 2.

21. A host cell containing a nucleic acid of Claim 20.

40



1/10

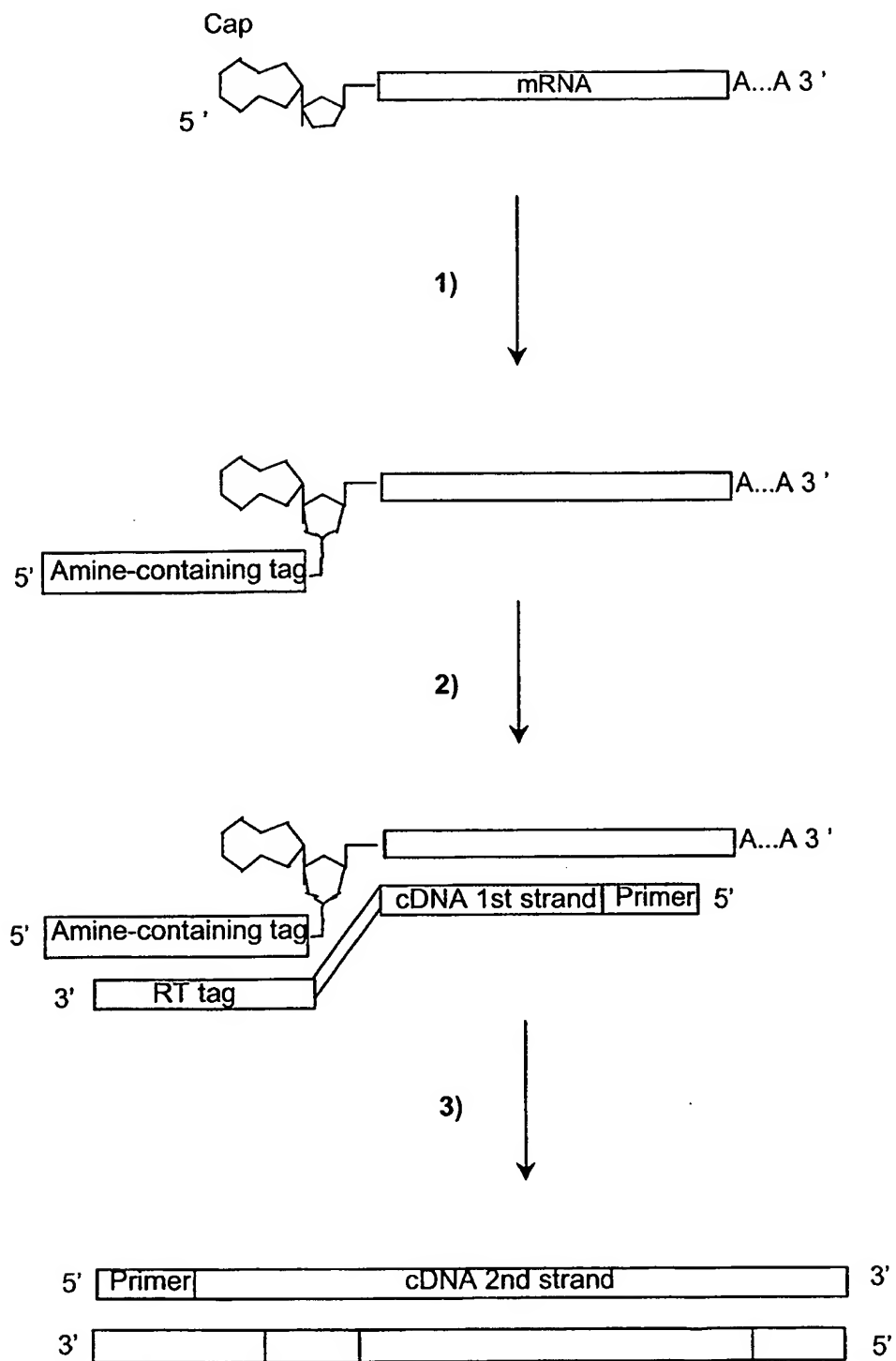


Figure 1

2/10

Minimum signal peptide score	false positive rate	false negative rate	proba(0.1)	proba(0.2)
3,5	0,121	0,036	0,467	0,664
4	0,096	0,06	0,519	0,708
4,5	0,078	0,079	0,565	0,745
5	0,062	0,098	0,615	0,782
5,5	0,05	0,127	0,659	0,813
6	0,04	0,163	0,694	0,836
6,5	0,033	0,202	0,725	0,855
7	0,025	0,248	0,763	0,878
7,5	0,021	0,304	0,78	0,889
8	0,015	0,368	0,816	0,909
8,5	0,012	0,418	0,836	0,92
9	0,009	0,512	0,856	0,93
9,5	0,007	0,581	0,863	0,934
10	0,006	0,679	0,835	0,919

Figure 2

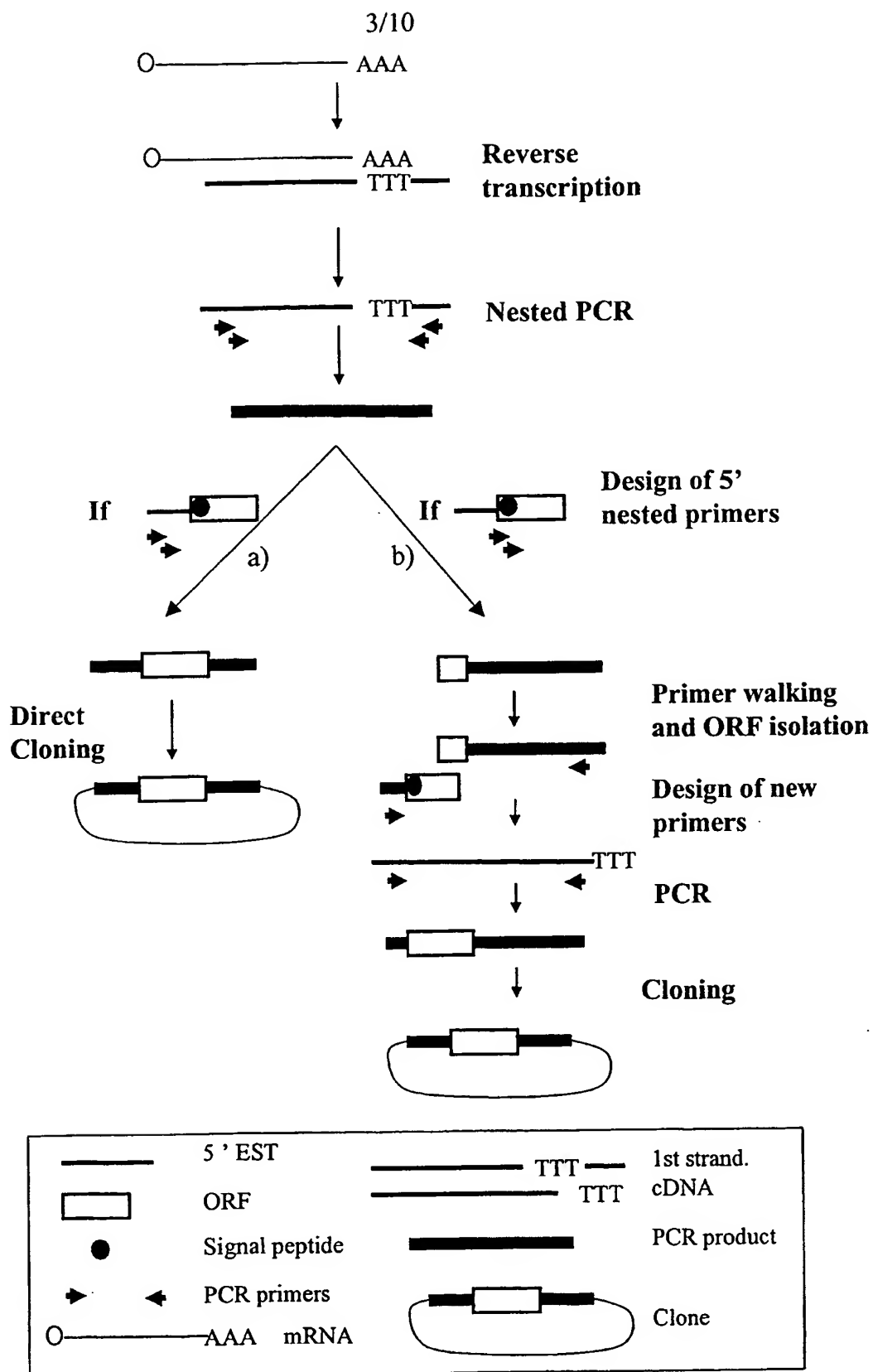


Figure 3

4/10

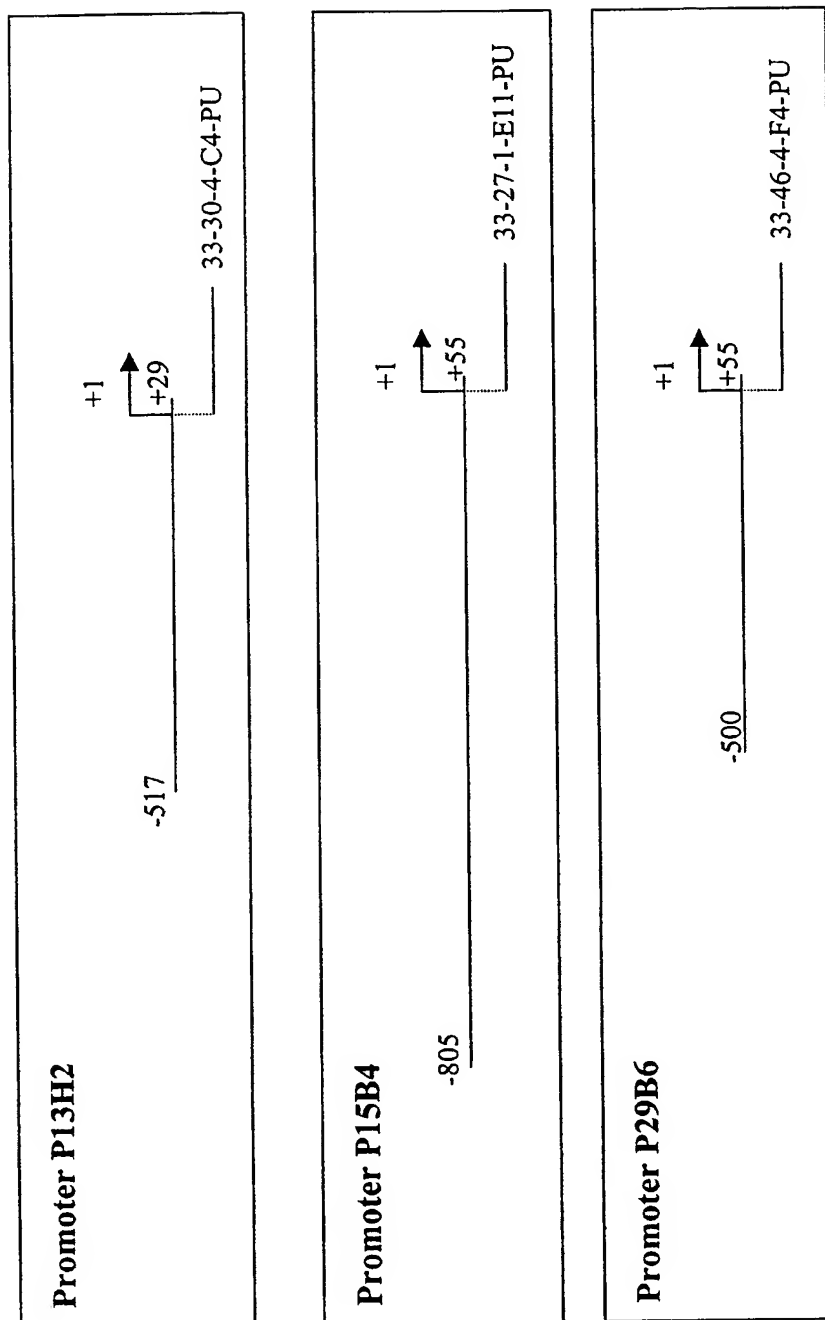


Figure 4

5/10

## Promoter sequence P13H2 (546 bp):

Matrix	Orient		Score	Length	Sequence
	Position	ation			
CMYB_01	-502	+	0.983	9	TGTCAGTTG
MYOD_Q6	-501	-	0.961	10	CCCAACTGAC
S8_01	-444	-	0.960	11	AATAGAATTAG
S8_01	-425	+	0.966	11	AACTAAATTAG
DELTAEF1_01	-390	-	0.960	11	GCACACCTCAG
GATA_C	-364	-	0.964	11	AGATAAATCCA
CMYB_01	-349	+	0.958	9	CTTCAGTTG
GATA1_02	-343	+	0.959	14	TTGTAGATAGGACA
GATA_C	-339	+	0.953	11	AGATAGGACAT
TAL1ALPHAE47_01	-235	+	0.973	16	CATAACAGATGGTAAG
TAL1BETAE47_01	-235	+	0.983	16	CATAACAGATGGTAAG
TAL1BETAITF2_01	-235	+	0.978	16	CATAACAGATGGTAAG
MYOD_Q6	-232	-	0.954	10	ACCATCTGTT
GATA1_04	-217	-	0.953	13	TCAAGATAAAGTA
IK1_01	-126	+	0.963	13	AGTTGGGAATTCC
IK2_01	-126	+	0.985	12	AGTTGGGAATTCC
CREL_01	-123	+	0.962	10	TGGGAATTCC
GATA1_02	-96	+	0.950	14	TCAGTGATATGGCA
SRY_02	-41	-	0.951	12	TAAAAACAAAACA
E2F_02	-33	+	0.957	8	TTAGCGC
MZF1_01	-5	-	0.975	8	TGAGGGGA

## Promoter sequence P15B4 (861bp) :

Matrix	Orient		Score	Length	Sequence
	Position	ation			
NFY_Q6	-748	-	0.956	11	GGACCAATCAT
MZF1_01	-738	+	0.962	8	CCTGGGGA
CMYB_01	-684	+	0.994	9	TGACCGTTG
VMYB_02	-682	-	0.985	9	TCCAACGGT
STAT_01	-673	+	0.968	9	TTCTTGGA
STAT_01	-673	-	0.951	9	TTCCAGGAA
MZF1_01	-556	-	0.956	8	TTGGGGGA
IK2_01	-451	+	0.965	12	GAATGGGATTTC
MZF1_01	-424	+	0.986	8	AGAGGGGA
SRY_02	-398	-	0.955	12	GAAAAACAAAACA
MZF1_01	-216	+	0.960	8	GAAGGGGA
MYOD_Q6	-190	+	0.981	10	AGCATCTGCC
DELTAEF1_01	-176	+	0.958	11	TCCCACCTTCC
S8_01	5	-	0.992	11	GAGGCAATTAT
MZF1_01	16	-	0.986	8	AGAGGGGA

## Promoter sequence P29B6 (555 bp) :

Matrix	Orient		Score	Length	Sequence
	Position	ation			
ARNT_01	-311	+	0.964	16	GGACTCACGTGCTGCT
NMYC_01	-309	+	0.965	12	ACTCACGTGCTG
USF_01	-309	+	0.985	12	ACTCACGTGCTG
USF_01	-309	-	0.985	12	CAGCACGTGAGT
NMYC_01	-309	-	0.956	12	CAGCACGTGAGT
MYCMAX_02	-309	-	0.972	12	CAGCACGTGAGT
USF_C	-307	+	0.997	8	TCACGTGC
USF_C	-307	-	0.991	8	GCACGTGA
MZF1_01	-292	-	0.968	8	CATGGGGA
ELK1_02	-105	+	0.963	14	CTCTCCGGAAGCCT
CETS1P54_01	-102	+	0.974	10	TCCGGAAGCC
AP1_Q4	-42	-	0.963	11	AGTGACTGAAC
AP1FJ_Q2	-42	-	0.961	11	AGTGACTGAAC
PADS_C	45	+	1.000	9	TGTGGTCTC

Figure 5

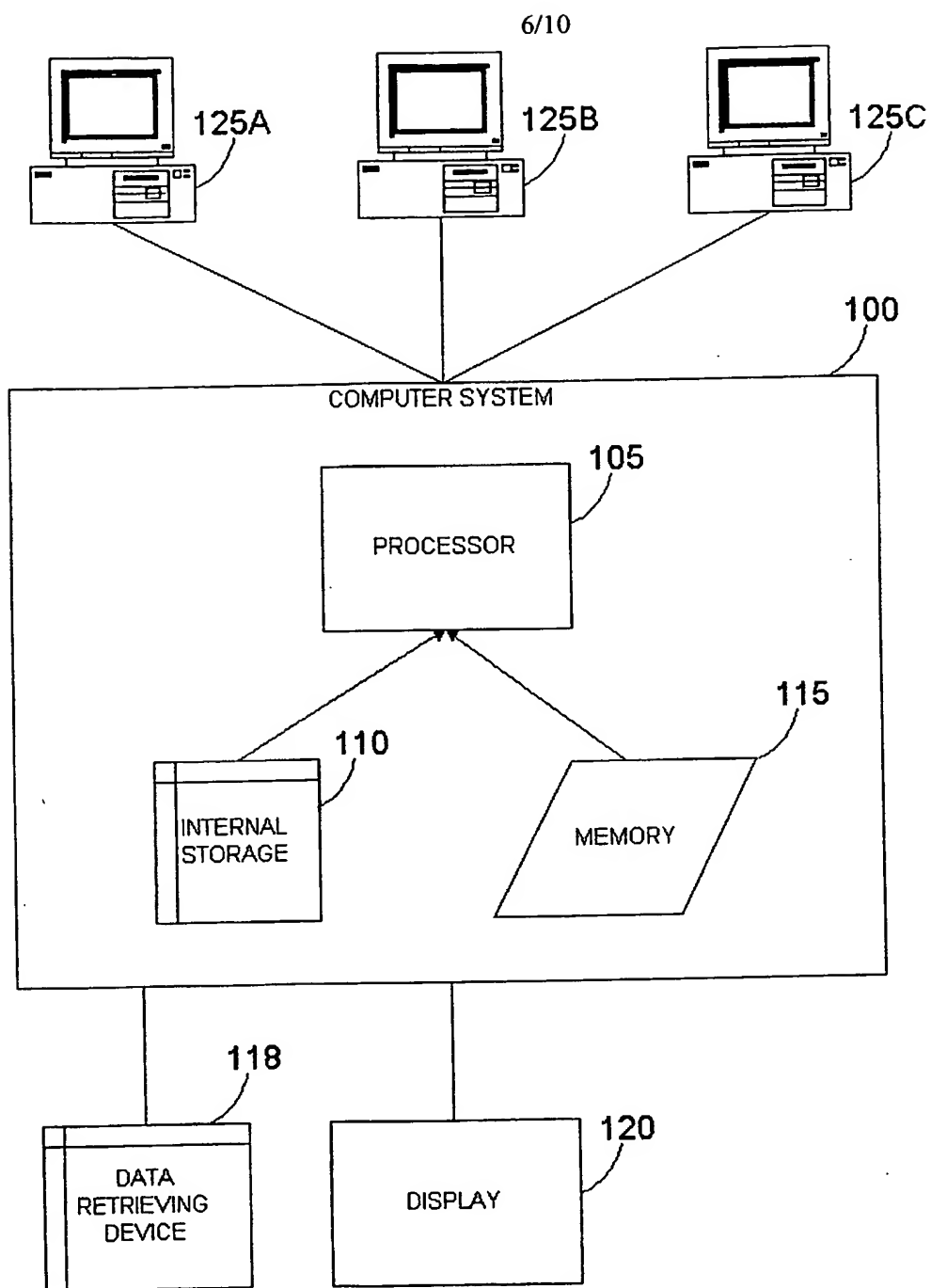


FIGURE 6

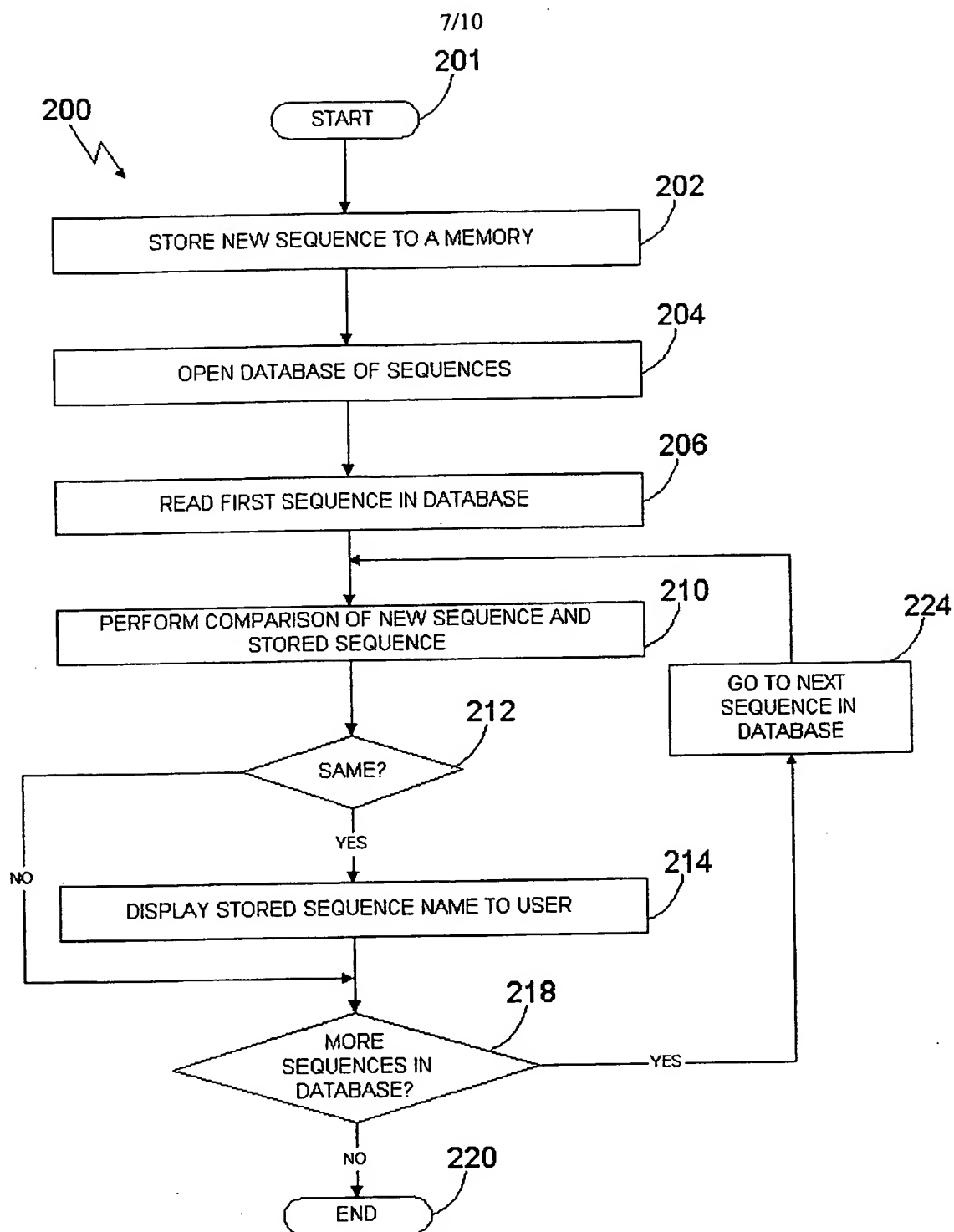


FIGURE 7

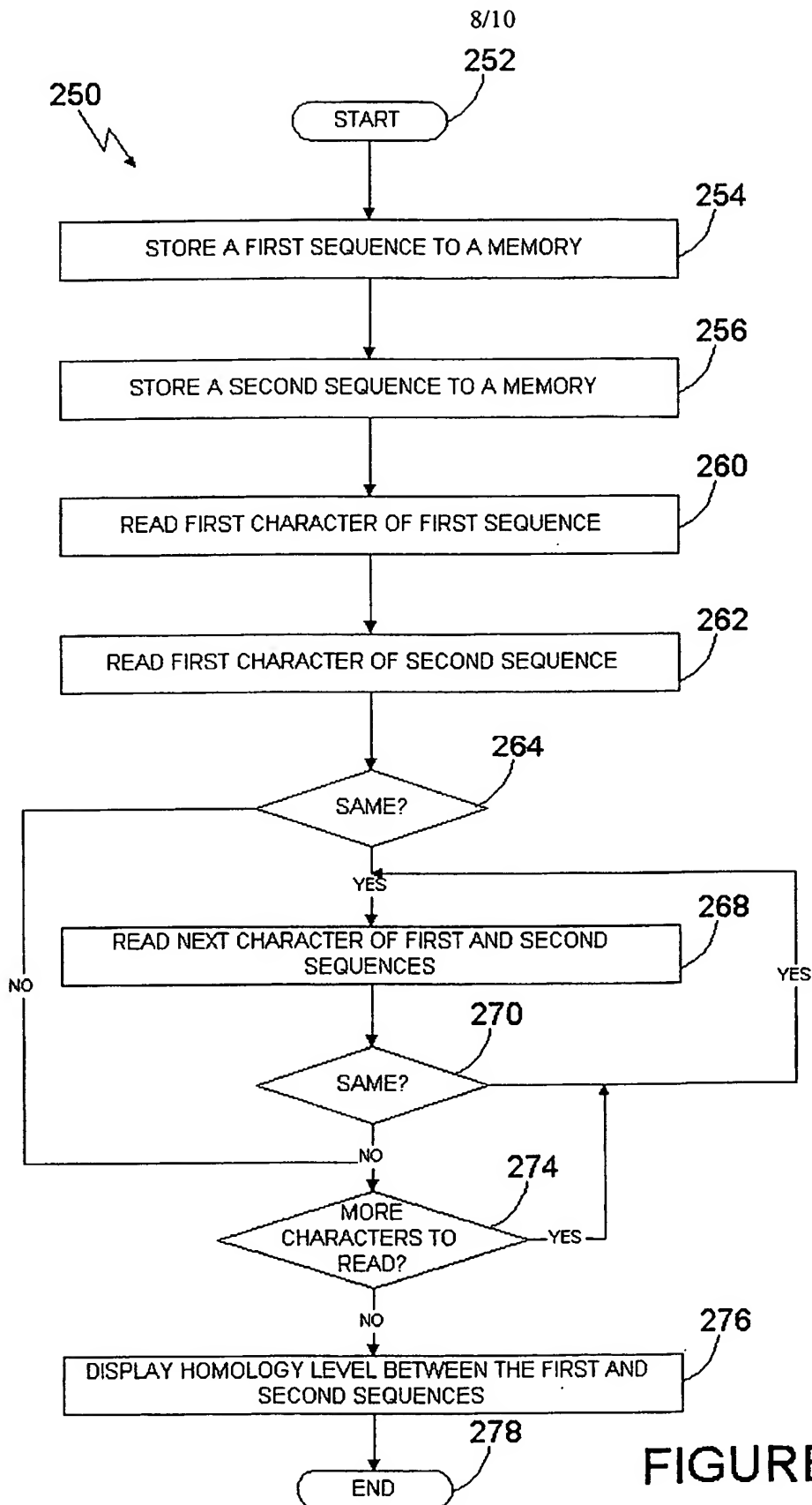


FIGURE 8



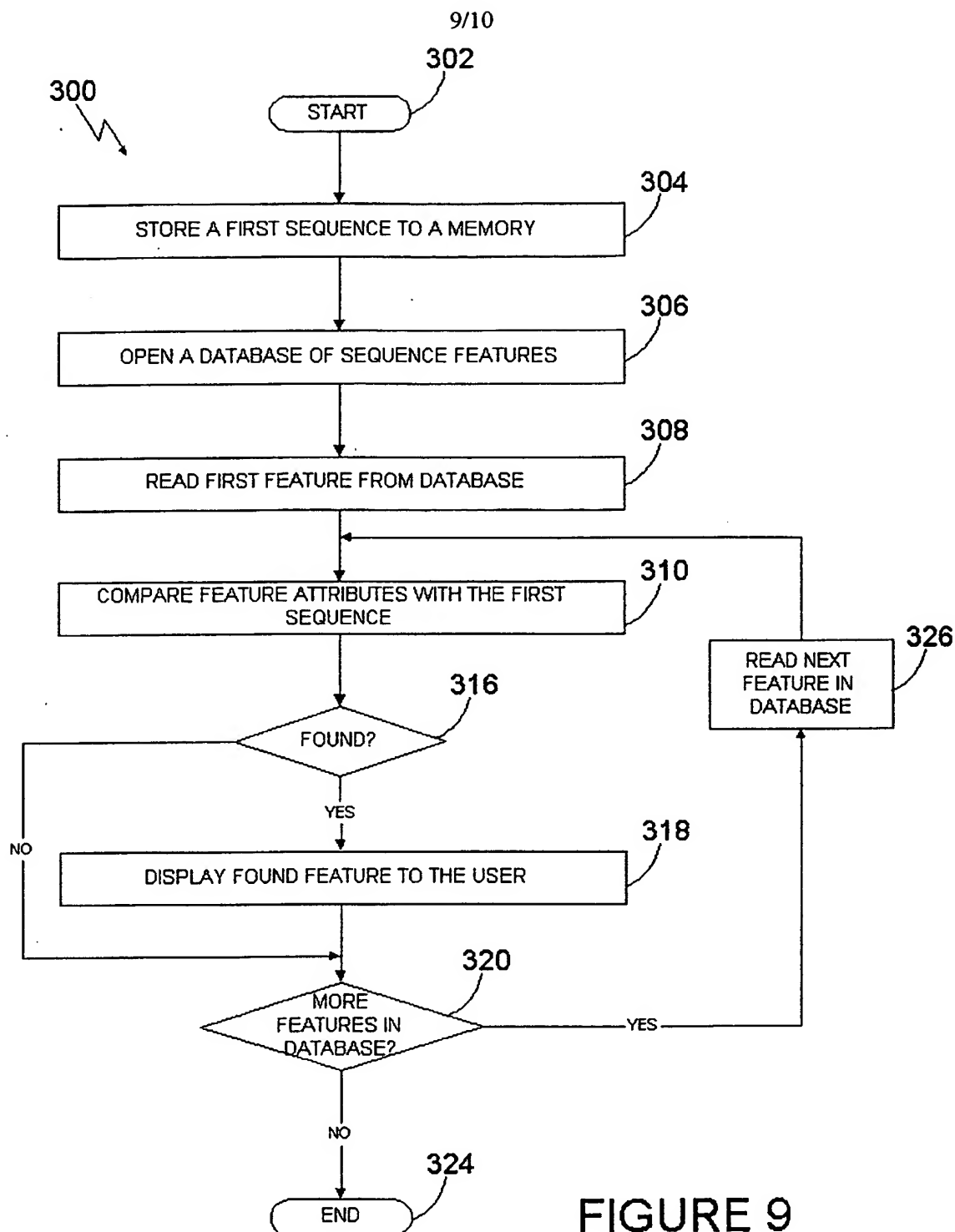


FIGURE 9

Step	Search characteristic		Selection Characteristics		
	Program	Strand	Parameters	Identity (%)	Length (bp)
miscellaneous	FASTA	both	-	90	15
tRNA	FASTA	both	-	80	60
rRNA	BLASTN	both	S=108	80	40
mtRNA	BLASTN	both	S=108	80	40
Prokaryotic	BLASTN	both	S=144	90	40
Fungal	BLASTN	both	S=144	90	40
Alu	BLASTN	both	S=72, B=5	70	40
L1	BLASTN	both	S=72, B=5	70	40
Repeats	BLASTN	both	S=72	70	40
PolyA	BLAST2N	top	W=6, S=10, E=1000, N=12	90	10
Polyadenylation signal	-	top	AATAAA allowing 1 mismatch		
Vertebrate	BLASTN then FASTA	both	-	90 then 70	30
ESTs	BLAST2N	both	-	90	30
Geneseq	BLASTN	both	W=8, B=10	90	30
ORF	BLASTP	top	W=8, B=10	-	-
Proteins	BLASTX	top	E = 0.001	70	30

Figure 10

## SEQUENCE LISTING

<110> Dumas Milne Edwards, J.B.  
 Duclert A.  
 Giordano, J.Y.  
 Genset SA

<120> ESTs and Encoded Human Proteins.

<130> D18118-339 881

<150> 09/057,719

<151> 1998-04-09

<150> 09/069,047

<151> 1998-04-28

<160> 1622

<170> Patent.pm

<210> 1

<211> 822

<212> DNA

<213> Homo Sapiens

<220>

<221> CDS

<222> 346..552

<221> sig\_peptide

<222> 346..408

<223> Von Heijne matrix

<221> misc\_feature

<222> 115

<223> n=a, g, c or t

<400> 1

```

actcctttta gcataggggc ttccggcgcca gcggccagcg ctagtcgggc tggtaagtgc      60
ctgatgccga gttccgtctc tcgcgtcttt tcctgggtccc aggcaaagcg gasgnagatc      120
ctcaaacggc ctagtgcttc gcgcttccgg agaaaatcag cggctctaatt aattcctctg      180
gtttgttgaa gcagttacca agaattctca accctttccc acaaaaagcta attgagtaca      240
cgttcctggt gagtacacgt tcctgttgat ttacaaaagg tgcaggtagt agcagggtctg      300
aagactaaca ttttgtgaag ttgtaaaaca gaaaacctgt tagaa atg tgg tgg ttt      357
                                     Met Trp Trp Phe
                                     -20
cag caa ggc ctc agt ttc ctt cct tca gcc ctt gta att tgg aca tct      405
Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val Ile Trp Thr Ser
      -15      -10      -5
gct gct ttc ata ttt tca tac att act gca gta aca ctc cac cat ata      453
Ala Ala Phe Ile Phe Ser Tyr Ile Thr Ala Val Thr Leu His His Ile
      1      5      10      15
gac ccg gct tta cct tat atc agt gac act ggt aca gta gct cca raa      501
Asp Pro Ala Leu Pro Tyr Ile Ser Asp Thr Gly Thr Val Ala Pro Xaa
      20      25      30
aaa tgc tta ttt ggg gca atg cta aat att gcg gca gtt tta tgt caa      549
Lys Cys Leu Phe Gly Ala Met Leu Asn Ile Ala Ala Val Leu Cys Gln
      35      40      45
aaa tagaaatcag gaarataatt caacttaaag aakttcattt catgaccaa      602
Lys
ctcttccaraa acatgtcttt acaagcatat ctcttgattt gctttctaca ctgttgaatt      662
gtctggcaat atttctgcag tggaaaattt gatttarmta gttcttgact gataaatatg      722

```

2

gtaaggtggg cttttccccc tgtgtaattg gctactatgt cttactgagc caagttgtaw 782  
 tttgaaataa aatgatatga gagtgacaca aaaaaaaaaa 822

<210> 2  
 <211> 21  
 <212> PRT  
 <213> Homo Sapiens

<220>  
 <221> SIGNAL  
 <222> 1..21  
 <223> Von Heijne matrix  
 score 5.5  
 seq SFLPSALVIWTS/AF

<400> 2  
 Met Trp Trp Phe Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val  
 1 5 10 15  
 Ile Trp Thr Ser Ala  
 20

<210> 3  
 <211> 526  
 <212> DNA  
 <213> Homo Sapiens

<220>  
 <221> CDS  
 <222> 90..344

<221> sig\_peptide  
 <222> 90..140  
 <223> Von Heijne matrix

<221> misc\_feature  
 <222> 290  
 <223> n=a, g, c or t

<400> 3  
 aatatrarac agctacaata ttccagggcc artcacttgc cattttctcat aacagcgtca 60  
 gagagaaaga actgactgar acgtttgag atg aag aaa gtt ctc ctc ctg atc 113  
 Met Lys Lys Val Leu Leu Leu Ile  
 -15 -10  
 aca gcc atc ttg gca gtg gct gtw ggt ttc cca gtc tct caa gac cag 161  
 Thr Ala Ile Leu Ala Val Ala Val Gly Phe Pro Val Ser Gln Asp Gln  
 -5 1 5  
 gaa cga gaa aaa aga agt atc agt gac agc gat gaa tta gct tca ggr 209  
 Glu Arg Glu Lys Arg Ser Ile Ser Asp Ser Asp Glu Leu Ala Ser Gly  
 10 15 20  
 wtt ttt gtg ttc cct tac cca tat cca ttt cgc cca ctt cca cca att 257  
 Xaa Phe Val Phe Pro Tyr Pro Tyr Pro Phe Arg Pro Leu Pro Pro Ile  
 25 30 35  
 cca ttt cca aga ttt cca tgg ttt aga cgt aan ttt cct att cca ata 305  
 Pro Phe Pro Arg Phe Pro Trp Phe Arg Arg Xaa Phe Pro Ile Pro Ile  
 40 45 50 55  
 cct gaa tct gcc cct aca act ccc ctt cct agc gaa aag taaacaaraa 354  
 Pro Glu Ser Ala Pro Thr Thr Pro Leu Pro Ser Glu Lys  
 60 65  
 ggaaaagtca crataaacct ggtcacctga aattgaaatt gagccacttc cttgaaraat 414  
 caaaattcct gttaataaaa raaaaacaaa tgtaattgaa atagcacaca gcatttctta 474  
 gtcaatatct ttagtgatct tctttaataa acatgaaagc aaaaaaaaaa aa 526

<210> 4

<211> 17  
 <212> PRT  
 <213> Homo Sapiens

<220>  
 <221> SIGNAL  
 <222> 1..17  
 <223> Von Heijne matrix  
 score 8.2  
 seq LLLITAILAVAVG/FP

<400> 4  
 Met Lys Lys Val Leu Leu Leu Ile Thr Ala Ile Leu Ala Val Ala Val  
 1 5 10 15  
 Gly

<210> 5  
 <211> 848  
 <212> DNA  
 <213> Homo Sapiens

<220>  
 <221> CDS  
 <222> 32..697

<221> sig\_peptide  
 <222> 32..73  
 <223> Von Heijne matrix

<400> 5  
 aactttgcct tgtgttttcc accctgaaag a atg ttg tgg ctg ctc ttt ttt 52  
 Met Leu Trp Leu Leu Phe Phe  
 -10  
 ctg gtg act gcc att cat gct gaa ctc tgt caa cca ggt gca gaa aat 100  
 Leu Val Thr Ala Ile His Ala Glu Leu Cys Gln Pro Gly Ala Glu Asn  
 -5 1 5  
 gct ttt aaa gtg aga ctt agt atc aga aca gct ctg gga gat aaa gca 148  
 Ala Phe Lys Val Arg Leu Ser Ile Arg Thr Ala Leu Gly Asp Lys Ala  
 10 15 20 25  
 tat gcc tgg gat acc aat gaa gaa tac ctc ttc aaa gcg atg gta gct 196  
 Tyr Ala Trp Asp Thr Asn Glu Glu Tyr Leu Phe Lys Ala Met Val Ala  
 30 35 40  
 ttc tcc atg aga aaa gtt ccc aac aga gaa gca aca gaa att tcc cat 244  
 Phe Ser Met Arg Lys Val Pro Asn Arg Glu Ala Thr Glu Ile Ser His  
 45 50 55  
 gtc cta ctt tgc aat gta acc cag agg gta tca ttc tgg ttt gtg gtt 292  
 Val Leu Leu Cys Asn Val Thr Gln Arg Val Ser Phe Trp Phe Val Val  
 60 65 70  
 aca gac cct tca aaa aat cac acc ctt cct gct gtt gag gtg caa tca 340  
 Thr Asp Pro Ser Lys Asn His Thr Leu Pro Ala Val Glu Val Gln Ser  
 75 80 85  
 gcc ata aga atg aac aag aac cgg atc aac aat gcc ttc ttt cta aat 388  
 Ala Ile Arg Met Asn Lys Asn Arg Ile Asn Asn Ala Phe Phe Leu Asn  
 90 95 100 105  
 gac caa act ctg gaa ttt tta aaa atc cct tcc aca ctt gca cca ccc 436  
 Asp Gln Thr Leu Glu Phe Leu Lys Ile Pro Ser Thr Leu Ala Pro Pro  
 110 115 120  
 atg gac cca tct gtg ccc atc tgg att att ata ttt ggt gtg ata ttt 484  
 Met Asp Pro Ser Val Pro Ile Trp Ile Ile Ile Phe Gly Val Ile Phe  
 125 130 135  
 tgc atc atc ata gtt gca att gca cta ctg att tta tca ggg atc tgg 532  
 Cys Ile Ile Ile Val Ala Ile Ala Leu Leu Ile Leu Ser Gly Ile Trp  
 140 145 150

```

caa cgt ada ara aag aac aaa gaa cca tct gaa gtg gat gac gct gaa      580
Gln Arg Xaa Xaa Lys Asn Lys Glu Pro Ser Glu Val Asp Asp Ala Glu
    155                160                165
rat aak tgt gaa aac atg atc aca att gaa aat ggc atc ccc tct gat      628
Xaa Xaa Cys Glu Asn Met Ile Thr Ile Glu Asn Gly Ile Pro Ser Asp
    170                175                180                185
ccc ctg gac atg aag gga ggg cat att aat gat gcc ttc atg aca gag      676
Pro Leu Asp Met Lys Gly Gly His Ile Asn Asp Ala Phe Met Thr Glu
                190                195                200
gat gag agg ctc acc cct ctc tgaagggtg ttgttctgct tcctcaaraa      727
Asp Glu Arg Leu Thr Pro Leu
                205
attaaacatt tgtttctgtg tgactgctga gcacctctgaa ataccaagag cagatcatat      787
wttttgtttc accattcttc ttttgtaata aattttgaaat gtgcttgaaa aaaaaaaaaa      847
c                                                                848

```

<210> 6  
 <211> 14  
 <212> PRT  
 <213> Homo Sapiens

<220>  
 <221> SIGNAL  
 <222> 1..14  
 <223> Von Heijne matrix  
 score 10.7  
 seq LWLLFPLVTAIHA/EL

<400> 6  
 Met Leu Trp Leu Leu Phe Phe Leu Val Thr Ala Ile His Ala  
 1 5 10

<210> 7  
 <211> 826  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 15..695

<221> sig\_peptide  
 <222> 15..80  
 <223> score 8.5  
 seq AALLLGLMMVVTG/DE

```

<400> 7
aaccagaggt gccc atg ggt tgg aca atg agg ctg gtc aca gca gca ctg      50
                Met Gly Trp Thr Met Arg Leu Val Thr Ala Ala Leu
                -20                -15
tta ctg ggt ctc atg atg gtg gtc act gga gac gag gat gag aac agc      98
Leu Leu Gly Leu Met Met Val Val Thr Gly Asp Glu Asp Glu Asn Ser
-10                -5                1                5
ccg tgt gcc cat gag gcc ctc ctg gac gag gac acc ctc ttt tgc cag      146
Pro Cys Ala His Glu Ala Leu Leu Asp Glu Asp Thr Leu Phe Cys Gln
                10                15                20
ggc ctt gaa gtt ttc tac cca gag ttg ggg aac att ggc tgc aag gtt      194
Gly Leu Glu Val Phe Tyr Pro Glu Leu Gly Asn Ile Gly Cys Lys Val
                25                30                35
gtt cct gat tgt aac aac tac aga cag aag atc acc tcc tgg atg gag      242
Val Pro Asp Cys Asn Asn Tyr Arg Gln Lys Ile Thr Ser Trp Met Glu
                40                45                50
ccg ata gtc aag ttc ccg ggg gcc gtg gac ggc gca acc tat atc ctg      290

```

```

Pro Ile Val Lys Phe Pro Gly Ala Val Asp Gly Ala Thr Tyr Ile Leu
55          60          65          70
gtg atg gtg gat cca gat gcc cct agc aga gca gaa ccc aga cag aga      338
Val Met Val Asp Pro Asp Ala Pro Ser Arg Ala Glu Pro Arg Gln Arg
          75          80          85
ttc tgg aga cat tgg ctg gta aca gat atc aag ggc gcc gac ctg aag      386
Phe Trp Arg His Trp Leu Val Thr Asp Ile Lys Gly Ala Asp Leu Lys
          90          95          100
aaa ggg aag att cag ggc cag gag tta tca gcc tac cag gct ccc tcc      434
Lys Gly Lys Ile Gln Gly Gln Glu Leu Ser Ala Tyr Gln Ala Pro Ser
          105          110          115
cca ccg gca cac agt ggc ttc cat cgc tac cag ttc ttt gtc tat ctt      482
Pro Pro Ala His Ser Gly Phe His Arg Tyr Gln Phe Phe Val Tyr Leu
          120          125          130
cag gaa gga aag gtc atc tct ctc ctt ccc aag gaa aac aaa act cga      530
Gln Glu Gly Lys Val Ile Ser Leu Leu Pro Lys Glu Asn Lys Thr Arg
          135          140          145          150
ggc tct tgg aaa atg gac aga ttt ctg aac cgt ttc cac ctg ggc gaa      578
Gly Ser Trp Lys Met Asp Arg Phe Leu Asn Arg Phe His Leu Gly Glu
          155          160          165
cct gaa gca agc acc cag ttc atg acc cag aac tac cag gac tca cca      626
Pro Glu Ala Ser Thr Gln Phe Met Thr Gln Asn Tyr Gln Asp Ser Pro
          170          175          180
acc ctc cag gct ccc aga gaa agg gcc agc gag ccc aag cac aaa aac      674
Thr Leu Gln Ala Pro Arg Glu Arg Ala Ser Glu Pro Lys His Lys Asn
          185          190          195
cag gcg gag ata gct gcc tgc tagatagccg gctttgccat ccgggcatgt      725
Gln Ala Glu Ile Ala Ala Cys
          200          205
ggccacactg cccaccaccg acgatgtggg tatggaaccc cctctggata cagaaccctt      785
tcttttccaa ataaaaaaaa aatcatccaa aaaaaaaaaa a      826

```

&lt;210&gt; 8

&lt;211&gt; 227

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SIGNAL

&lt;222&gt; -22...-1

&lt;223&gt; score 8.5

seq AALLLGLMMVVTG/DE

&lt;400&gt; 8

```

Met Gly Trp Thr Met Arg Leu Val Thr Ala Ala Leu Leu Leu Gly Leu
          -20          -15          -10
Met Met Val Val Thr Gly Asp Glu Asp Glu Asn Ser Pro Cys Ala His
          -5          1          5          10
Glu Ala Leu Leu Asp Glu Asp Thr Leu Phe Cys Gln Gly Leu Glu Val
          15          20          25
Phe Tyr Pro Glu Leu Gly Asn Ile Gly Cys Lys Val Val Pro Asp Cys
          30          35          40
Asn Asn Tyr Arg Gln Lys Ile Thr Ser Trp Met Glu Pro Ile Val Lys
          45          50          55
Phe Pro Gly Ala Val Asp Gly Ala Thr Tyr Ile Leu Val Met Val Asp
          60          65          70
Pro Asp Ala Pro Ser Arg Ala Glu Pro Arg Gln Arg Phe Trp Arg His
          75          80          85          90
Trp Leu Val Thr Asp Ile Lys Gly Ala Asp Leu Lys Lys Gly Lys Ile
          95          100          105
Gln Gly Gln Glu Leu Ser Ala Tyr Gln Ala Pro Ser Pro Pro Ala His
          110          115          120
Ser Gly Phe His Arg Tyr Gln Phe Phe Val Tyr Leu Gln Glu Gly Lys

```

```

      125      130      135
Val Ile Ser Leu Leu Pro Lys Glu Asn Lys Thr Arg Gly Ser Trp Lys
      140      145      150
Met Asp Arg Phe Leu Asn Arg Phe His Leu Gly Glu Pro Glu Ala Ser
155      160      165      170
Thr Gln Phe Met Thr Gln Asn Tyr Gln Asp Ser Pro Thr Leu Gln Ala
      175      180      185
Pro Arg Glu Arg Ala Ser Glu Pro Lys His Lys Asn Gln Ala Glu Ile
      190      195      200
Ala Ala Cys
      205

```

<210> 9  
 <211> 852  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 229..735

<221> sig\_peptide  
 <222> 229..492  
 <223> score 6.7  
 seq VFALSSFLNKASA/VY

```

<400> 9
aatgactggc agtggcatca gcgatggcgg ctgcgctcggg gtcgggttctg cagcgtgtga      60
tcgtgtcgcc ggcagggagg catagcgcc ctctgatctt cctgcatggc tcaggtgatt      120
ctggacaagg attaagaatg tggatcaagc aggttttttaa atcaagattt aacattccaa      180
cacataaaaa ttatttatcc aacagctcct cccagatcat atactcct atg aaa gga      237
                                     Met Lys Gly
gga atc tcc aat gta tgg ttt gac aga ttt aaa ata acc aat gac tgc      285
Gly Ile Ser Asn Val Trp Phe Asp Arg Phe Lys Ile Thr Asn Asp Cys
-85      -80      -75      -70
cca gaa cac ctt gaa tca att gat gtc atg tgt caa gtg ctt act gat      333
Pro Glu His Leu Glu Ser Ile Asp Val Met Cys Gln Val Leu Thr Asp
      -65      -60      -55
ttg att gat gaa gaa gta aaa agt ggc atc aag aag aac agg ata tta      381
Leu Ile Asp Glu Glu Val Lys Ser Gly Ile Lys Lys Asn Arg Ile Leu
      -50      -45      -40
ata gga gga ttc tct atg gga gga tgc atg gca atg cat tta gca tat      429
Ile Gly Gly Phe Ser Met Gly Gly Cys Met Ala Met His Leu Ala Tyr
      -35      -30      -25
aga aat cat caa gat gtg gca gga gta ttt gct ctt tct agt ttt ctg      477
Arg Asn His Gln Asp Val Ala Gly Val Phe Ala Leu Ser Ser Phe Leu
      -20      -15      -10
aat aaa gca tct gct gtt tac cag gct ctt cag aag agt aat ggt gta      525
Asn Lys Ala Ser Ala Val Tyr Gln Ala Leu Gln Lys Ser Asn Gly Val
      -5      1      5      10
ctt cct gaa tta ttt cag tgt cat ggt act gca gat gag tta gtt ctt      573
Leu Pro Glu Leu Phe Gln Cys His Gly Thr Ala Asp Glu Leu Val Leu
      15      20      25
cat tct tgg gca gaa gag aca aac tca atg tta aaa tct cta gga gtg      621
His Ser Trp Ala Glu Glu Thr Asn Ser Met Leu Lys Ser Leu Gly Val
      30      35      40
acc acg aag ttt cat agt ttt cca aat gtt tac cat gag cta agc aaa      669
Thr Thr Lys Phe His Ser Phe Pro Asn Val Tyr His Glu Leu Ser Lys
      45      50      55
act gag tta gac ata ttg aag tta tgg att ctt aca aag ctg cca gga      717
Thr Glu Leu Asp Ile Leu Lys Leu Trp Ile Leu Thr Lys Leu Pro Gly
      60      65      70      75
gaa atg gaa aaa caa aaa tgaatgaatc aagagtgatt tgttaatgta      765

```



Glu Met Glu Lys Gln Lys

80

agtgtaatgt ctttgtgaaa agtgattttt actgccaaat tataatgata attaaaaatat 825  
 taagaaatag caaaaaaaaa aaaaaaa 852

&lt;210&gt; 10

&lt;211&gt; 169

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SIGNAL

&lt;222&gt; -88..-1

<223> score 6.7  
 seq VFALSSFLNKASA/VY

&lt;400&gt; 10

Met Lys Gly Gly Ile Ser Asn Val Trp Phe Asp Arg Phe Lys Ile Thr  
 -85 -80 -75  
 Asn Asp Cys Pro Glu His Leu Glu Ser Ile Asp Val Met Cys Gln Val  
 -70 -65 -60  
 Leu Thr Asp Leu Ile Asp Glu Glu Val Lys Ser Gly Ile Lys Lys Asn  
 -55 -50 -45  
 Arg Ile Leu Ile Gly Gly Phe Ser Met Gly Gly Cys Met Ala Met His  
 -40 -35 -30 -25  
 Leu Ala Tyr Arg Asn His Gln Asp Val Ala Gly Val Phe Ala Leu Ser  
 -20 -15 -10  
 Ser Phe Leu Asn Lys Ala Ser Ala Val Tyr Gln Ala Leu Gln Lys Ser  
 -5 1 5  
 Asn Gly Val Leu Pro Glu Leu Phe Gln Cys His Gly Thr Ala Asp Glu  
 10 15 20  
 Leu Val Leu His Ser Trp Ala Glu Glu Thr Asn Ser Met Leu Lys Ser  
 25 30 35 40  
 Leu Gly Val Thr Thr Lys Phe His Ser Phe Pro Asn Val Tyr His Glu  
 45 50 55  
 Leu Ser Lys Thr Glu Leu Asp Ile Leu Lys Leu Trp Ile Leu Thr Lys  
 60 65 70  
 Leu Pro Gly Glu Met Glu Lys Gln Lys  
 75 80

&lt;210&gt; 11

&lt;211&gt; 1602

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 24..1004

&lt;221&gt; sig\_peptide

&lt;222&gt; 24..170

<223> score 5.6  
 seq ACLSLGFFSLLWL/QL

&lt;400&gt; 11

atgcgcgcc gcctctccgc acg atg ttc ccc tcg cgg agg aaa gcg gcg cag 53  
 Met Phe Pro Ser Arg Arg Lys Ala Ala Gln  
 -45 -40  
 ctg ccc tgg gag gac ggc agg tcc ggg ttg ctc tcc ggc ggc ctc cct 101  
 Leu Pro Trp Glu Asp Gly Arg Ser Gly Leu Leu Ser Gly Gly Leu Pro  
 -35 -30 -25  
 cgg aag tgt tcc gtc ttc cac ctg ttc gtg gcc tgc ctc tcg ctg ggc 149  
 Arg Lys Cys Ser Val Phe His Leu Phe Val Ala Cys Leu Ser Leu Gly



```
<210> 12
<211> 327
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -49..-1
<223> score 5.6
      seq ACLSLGFFSLLWL/OL
```

<400>	12																
Met	Phe	Pro	Ser	Arg	Arg	Lys	Ala	Ala	Gln	Leu	Pro	Trp	Glu	Asp	Gly		
				-45					-40					-35			
Arg	Ser	Gly	Leu	Leu	Ser	Gly	Gly	Leu	Pro	Arg	Lys	Cys	Ser	Val	Phe		
			-30					-25					-20				
His	Leu	Phe	Val	Ala	Cys	Leu	Ser	Leu	Gly	Phe	Phe	Ser	Leu	Leu	Trp		
		-15						-10				-5					
Leu	Gln	Leu	Ser	Cys	Ser	Gly	Asp	Val	Ala	Arg	Ala	Val	Arg	Gly	Gln		
	1				5					10					15		
Gly	Gln	Glu	Thr	Ser	Gly	Pro	Pro	Arg	Ala	Cys	Pro	Pro	Glu	Pro	Pro		
			20						25					30			
Pro	Glu	His	Trp	Glu	Glu	Asp	Ala	Ser	Trp	Gly	Pro	His	Arg	Leu	Ala		
			35					40					45				
Val	Leu	Val	Pro	Phe	Arg	Glu	Arg	Phe	Glu	Glu	Leu	Leu	Val	Phe	Val		
		50					55					60					
Pro	His	Met	Arg	Arg	Phe	Leu	Ser	Arg	Lys	Lys	Ile	Arg	His	His	Ile		
	65					70					75						
Tyr	Val	Leu	Asn	Gln	Val	Asp	His	Phe	Arg	Phe	Asn	Arg	Ala	Ala	Leu		
80					85					90					95		
Ile	Asn	Val	Gly	Phe	Leu	Glu	Ser	Ser	Asn	Ser	Thr	Asp	Tyr	Ile	Ala		
				100					105					110			
Met	His	Asp	Val	Asp	Leu	Leu	Pro	Leu	Asn	Glu	Glu	Leu	Asp	Tyr	Gly		
			115					120					125				
Phe	Pro	Glu	Ala	Gly	Pro	Phe	His	Val	Ala	Ser	Pro	Glu	Leu	His	Pro		
		130					135					140					
Leu	Tyr	His	Tyr	Lys	Thr	Tyr	Val	Gly	Gly	Ile	Leu	Leu	Leu	Ser	Lys		
	145					150					155						
Gln	His	Tyr	Arg	Leu	Cys	Asn	Gly	Met	Ser	Asn	Arg	Phe	Trp	Gly	Trp		
160					165					170					175		
Gly	Arg	Glu	Asp	Asp	Glu	Phe	Tyr	Arg	Arg	Ile	Lys	Gly	Ala	Gly	Leu		
			180						185						190		
Gln	Leu	Phe	Arg	Pro	Ser	Gly	Ile	Thr	Thr	Gly	Tyr	Lys	Thr	Phe	Arg		
			195					200					205				
His	Leu	His	Asp	Pro	Ala	Trp	Arg	Lys	Arg	Asp	Gln	Lys	Arg	Ile	Ala		
		210					215					220					
Ala	Gln	Lys	Gln	Glu	Gln	Phe	Lys	Val	Asp	Arg	Glu	Gly	Gly	Leu	Asn		
		225				230					235						
Thr	Val	Lys	Tyr	His	Val	Ala	Ser	Arg	Thr	Ala	Leu	Ser	Val	Gly	Gly		
240				245						250					255		
Ala	Pro	Cys	Thr	Val	Leu	Asn	Ile	Met	Leu	Asp	Cys	Asp	Lys	Thr	Ala		
			260						265			</					

```
<210> 13
<211> 1568
<212> DNA
<213> Homo sapiens
```

<220>  
 <221> CDS  
 <222> 75..1259

<221> sig\_peptide  
 <222> 75..1004  
 <223> score 4.4  
 seq VLILLFSLALIIL/PS

<400> 13  
 agaaaagggtg tagtgtttgg ggcgggtcaac gggctatgct ggcttgacag ggctgggctc 60  
 ttcagaacag aagc atg gat ctc gga atc cct gac ctg ctg gac gcg tgg 110  
 Met Asp Leu Gly Ile Pro Asp Leu Leu Asp Ala Trp  
 -310 -305 -300  
 ctg gag ccc cca gag gat atc ttc tcg aca gga tcc gtc ctg gag ctg 158  
 Leu Glu Pro Pro Glu Asp Ile Phe Ser Thr Gly Ser Val Leu Glu Leu  
 -295 -290 -285  
 gga ctc cac tgc ccc cct cca gag gtt ccg gta act agg cta cag gaa 206  
 Gly Leu His Cys Pro Pro Pro Glu Val Pro Val Thr Arg Leu Gln Glu  
 -280 -275 -270  
 cag gga ctg caa ggc tgg aag tcc ggt ggg gac cgt ggc tgt ggc ctt 254  
 Gln Gly Leu Gln Gly Trp Lys Ser Gly Gly Asp Arg Gly Cys Gly Leu  
 -265 -260 -255  
 caa gag agt gag cct gaa gat ttc ttg aag ctt ttc att gat ccc aat 302  
 Gln Glu Ser Glu Pro Glu Asp Phe Leu Lys Leu Phe Ile Asp Pro Asn  
 -250 -245 -240 -235  
 gag gtg tac tgc tca gaa gca tct cct ggc agt gac agt ggc atc tct 350  
 Glu Val Tyr Cys Ser Glu Ala Ser Pro Gly Ser Asp Ser Gly Ile Ser  
 -230 -225 -220  
 gag gac tcc tgc cat cca gac agt ccc cct gcc ccc agg gca acc agt 398  
 Glu Asp Ser Cys His Pro Asp Ser Pro Pro Ala Pro Arg Ala Thr Ser  
 -215 -210 -205  
 tct cct atg ctc tat gag gtt gtc tat gag gca ggg gcc ctg gag agg 446  
 Ser Pro Met Leu Tyr Glu Val Val Tyr Glu Ala Gly Ala Leu Glu Arg  
 -200 -195 -190  
 atg cag ggg gaa act ggg cca aat gta ggc ctt atc tcc atc cag cta 494  
 Met Gln Gly Glu Thr Gly Pro Asn Val Gly Leu Ile Ser Ile Gln Leu  
 -185 -180 -175  
 gat cag tgg agc cca gca ttt atg gtg cct gat tcc tgc atg gtc agt 542  
 Asp Gln Trp Ser Pro Ala Phe Met Val Pro Asp Ser Cys Met Val Ser  
 -170 -165 -160 -155  
 gag ctg ccc ttt gat gct cat gcc cac atc ctg ccc aga gca ggc acc 590  
 Glu Leu Pro Phe Asp Ala His Ala His Ile Leu Pro Arg Ala Gly Thr  
 -150 -145 -140  
 gta gcc cca gtg ccc tgt aca acc ctg ctg ccc tgt caa acc ctg ttc 638  
 Val Ala Pro Val Pro Cys Thr Thr Leu Leu Pro Cys Gln Thr Leu Phe  
 -135 -130 -125  
 ctg acc gat gag gag aag cgt ctg ctg ggg cag gaa ggg gtt tcc ctg 686  
 Leu Thr Asp Glu Glu Lys Arg Leu Leu Gly Gln Glu Gly Val Ser Leu  
 -120 -115 -110  
 ccc tct cac ctg ccc ctc acc aag gca gag gag agg gtc ctc aag aag 734  
 Pro Ser His Leu Pro Leu Thr Lys Ala Glu Glu Arg Val Leu Lys Lys  
 -105 -100 -95  
 gtc agg agg aaa atc cgt aac aag cag tca gct cag gac agt cgg cgg 782  
 Val Arg Arg Lys Ile Arg Asn Lys Gln Ser Ala Gln Asp Ser Arg Arg  
 -90 -85 -80 -75  
 cgg aag aag gag tac att gat ggg ctg gag agc agg gtg gca gcc tgt 830  
 Arg Lys Lys Glu Tyr Ile Asp Gly Leu Glu Ser Arg Val Ala Ala Cys  
 -70 -65 -60  
 tct gca cag aac caa gaa tta cag aaa aaa gtc cag gag ctg gag agg 878  
 Ser Ala Gln Asn Gln Glu Leu Gln Lys Lys Val Gln Glu Leu Glu Arg  
 -55 -50 -45  
 cac aac atc tcc ttg gta gct cag ctc cgc cag ctg cag acg cta att 926

```

His Asn Ile Ser Leu Val Ala Gln Leu Arg Gln Leu Gln Thr Leu Ile
      -40          -35          -30
gct caa act tcc aac aaa gct gcc cag acc agc act tgt gtt ttg att      974
Ala Gln Thr Ser Asn Lys Ala Ala Gln Thr Ser Thr Cys Val Leu Ile
      -25          -20          -15
ctt ctt ttt tcc ctg gct ctc atc atc ctg ccc agc ttc agt cca ttc      1022
Leu Leu Phe Ser Leu Ala Leu Ile Ile Leu Pro Ser Phe Ser Pro Phe
      -10          -5          1          5
cag agt cga cca gaa gct ggg tct gag gat tac cag cct cac gga gtg      1070
Gln Ser Arg Pro Glu Ala Gly Ser Glu Asp Tyr Gln Pro His Gly Val
      10          15          20
act tcc aga aat atc ctg acc cac aag gac gta aca gaa aat ctg gag      1118
Thr Ser Arg Asn Ile Leu Thr His Lys Asp Val Thr Glu Asn Leu Glu
      25          30          35
acc caa gtg gta gag tcc aga ctg agg gag cca cct gga gcc aag gat      1166
Thr Gln Val Val Glu Ser Arg Leu Arg Glu Pro Pro Gly Ala Lys Asp
      40          45          50
gca aat ggc tca aca agg aca ctg ctt gag aag atg gga ggg aag cca      1214
Ala Asn Gly Ser Thr Arg Thr Leu Leu Glu Lys Met Gly Gly Lys Pro
      55          60          65          70
aga ccc agt ggg cgc atc cgg tcc gtg ctg cat gca gat gag atg      1259
Arg Pro Ser Gly Arg Ile Arg Ser Val Leu His Ala Asp Glu Met
      75          80          85
tgagctggaa cagaccttcc tggcccactt cctgatcaca aggaatcctg ggcttcctta      1319
tggctttctt cccactggga ttctactta ggtgtctgcc ctccagggtc caaatcactt      1379
caggacaccc caagagatgt cctttagtct ctgcctgagg cctagtctgc attgttttgc      1439
atatatgaga gggtagctca aatacttctg ttatgtatct gtgattttat ttcttctttg      1499
ggtatagggt tgaggggaaa taagttttga gtgagaaata aacgttttag ctgaaaaaaa      1559
aaaaaaaaa

```

&lt;210&gt; 14

&lt;211&gt; 395

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SIGNAL

&lt;222&gt; -310...-1

&lt;223&gt; score 4.4

seq VLILLFSLALIIL/PS

&lt;400&gt; 14

```

Met Asp Leu Gly Ile Pro Asp Leu Leu Asp Ala Trp Leu Glu Pro Pro
-310          -305          -300          -295
Glu Asp Ile Phe Ser Thr Gly Ser Val Leu Glu Leu Gly Leu His Cys
      -290          -285          -280
Pro Pro Pro Glu Val Pro Val Thr Arg Leu Gln Glu Gln Gly Leu Gln
      -275          -270          -265
Gly Trp Lys Ser Gly Gly Asp Arg Gly Cys Gly Leu Gln Glu Ser Glu
      -260          -255          -250
Pro Glu Asp Phe Leu Lys Leu Phe Ile Asp Pro Asn Glu Val Tyr Cys
      -245          -240          -235
Ser Glu Ala Ser Pro Gly Ser Asp Ser Gly Ile Ser Glu Asp Ser Cys
      -230          -225          -220          -215
His Pro Asp Ser Pro Pro Ala Pro Arg Ala Thr Ser Ser Pro Met Leu
      -210          -205          -200
Tyr Glu Val Val Tyr Glu Ala Gly Ala Leu Glu Arg Met Gln Gly Glu
      -195          -190          -185
Thr Gly Pro Asn Val Gly Leu Ile Ser Ile Gln Leu Asp Gln Trp Ser
      -180          -175          -170
Pro Ala Phe Met Val Pro Asp Ser Cys Met Val Ser Glu Leu Pro Phe
      -165          -160          -155
Asp Ala His Ala His Ile Leu Pro Arg Ala Gly Thr Val Ala Pro Val

```

12

-150                      -145                      -140                      -135  
 Pro Cys Thr Thr Leu Leu Pro Cys Gln Thr Leu Phe Leu Thr Asp Glu  
                                  -130                      -125                      -120  
 Glu Lys Arg Leu Leu Gly Gln Glu Gly Val Ser Leu Pro Ser His Leu  
                                  -115                      -110                      -105  
 Pro Leu Thr Lys Ala Glu Glu Arg Val Leu Lys Lys Val Arg Arg Lys  
                                  -100                      -95                      -90  
 Ile Arg Asn Lys Gln Ser Ala Gln Asp Ser Arg Arg Arg Lys Lys Glu  
                                  -85                      -80                      -75  
 Tyr Ile Asp Gly Leu Glu Ser Arg Val Ala Ala Cys Ser Ala Gln Asn  
                                  -70                      -65                      -60                      -55  
 Gln Glu Leu Gln Lys Lys Val Gln Glu Leu Glu Arg His Asn Ile Ser  
                                  -50                      -45                      -40  
 Leu Val Ala Gln Leu Arg Gln Leu Gln Thr Leu Ile Ala Gln Thr Ser  
                                  -35                      -30                      -25  
 Asn Lys Ala Ala Gln Thr Ser Thr Cys Val Leu Ile Leu Leu Phe Ser  
                                  -20                      -15                      -10  
 Leu Ala Leu Ile Ile Leu Pro Ser Phe Ser Pro Phe Gln Ser Arg Pro  
                                  -5                      1                      5                      10  
 Glu Ala Gly Ser Glu Asp Tyr Gln Pro His Gly Val Thr Ser Arg Asn  
                                  15                      20                      25  
 Ile Leu Thr His Lys Asp Val Thr Glu Asn Leu Glu Thr Gln Val Val  
                                  30                      35                      40  
 Glu Ser Arg Leu Arg Glu Pro Pro Gly Ala Lys Asp Ala Asn Gly Ser  
                                  45                      50                      55  
 Thr Arg Thr Leu Leu Glu Lys Met Gly Gly Lys Pro Arg Pro Ser Gly  
                                  60                      65                      70  
 Arg Ile Arg Ser Val Leu His Ala Asp Glu Met  
 75                      80                      85

&lt;210&gt; 15

&lt;211&gt; 25

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;400&gt; 15

gggaagatgg agatagtatt gcctg

25

&lt;210&gt; 16

&lt;211&gt; 26

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;400&gt; 16

ctgccatgta catgatagag agattc

26

&lt;210&gt; 17

&lt;211&gt; 546

&lt;212&gt; DNA

&lt;213&gt; Homo Sapiens

&lt;220&gt;

&lt;221&gt; promoter

&lt;222&gt; 1..517

&lt;221&gt; transcription start site

&lt;222&gt; 518

&lt;221&gt; protein\_bind

&lt;222&gt; 17..25

&lt;223&gt; matinspector prediction

name CMYB\_01

score 0.983

sequence tgtcagttg

<221> protein\_bind  
<222> complement(18..27)  
<223> matinspector prediction  
name MYOD\_Q6  
score 0.961  
sequence cccaactgac

<221> protein\_bind  
<222> complement(75..85)  
<223> matinspector prediction  
name S8\_01  
score 0.960  
sequence aatagaattag

<221> protein\_bind  
<222> 94..104  
<223> matinspector prediction  
name S8\_01  
score 0.966  
sequence aactaaattag

<221> protein\_bind  
<222> complement(129..139)  
<223> matinspector prediction  
name DELTAEF1\_01  
score 0.960  
sequence gcacacctcag

<221> protein\_bind  
<222> complement(155..165)  
<223> matinspector prediction  
name GATA\_C  
score 0.964  
sequence agataaatcca

<221> protein\_bind  
<222> 170..178  
<223> matinspector prediction  
name CMYB\_01  
score 0.958  
sequence cttcagttg

<221> protein\_bind  
<222> 176..189  
<223> matinspector prediction  
name GATA1\_02  
score 0.959  
sequence ttgtagataggaca

<221> protein\_bind  
<222> 180..190  
<223> matinspector prediction  
name GATA\_C  
score 0.953  
sequence agataggacat

<221> protein\_bind  
<222> 284..299  
<223> matinspector prediction  
name TAL1ALPHA47\_01  
score 0.973  
sequence cataacagatggtaag

<221> protein\_bind  
<222> 284..299  
<223> matinspector prediction  
name TAL1BETAE47\_01  
score 0.983  
sequence cataacagatggtaag

<221> protein\_bind  
<222> 284..299  
<223> matinspector prediction  
name TAL1BETAITF2\_01  
score 0.978  
sequence cataacagatggtaag

<221> protein\_bind  
<222> complement(287..296)  
<223> matinspector prediction  
name MYOD\_Q6  
score 0.954  
sequence accatctggt

<221> protein\_bind  
<222> complement(302..314)  
<223> matinspector prediction  
name GATA1\_04  
score 0.953  
sequence tcaagataaagta

<221> protein\_bind  
<222> 393..405  
<223> matinspector prediction  
name IK1\_01  
score 0.963  
sequence agttgggaattcc

<221> protein\_bind  
<222> 393..404  
<223> matinspector prediction  
name IK2\_01  
score 0.985  
sequence agttgggaattc

<221> protein\_bind  
<222> 396..405  
<223> matinspector prediction  
name CREL\_01  
score 0.962  
sequence tgggaattcc

<221> protein\_bind  
<222> 423..436  
<223> matinspector prediction  
name GATA1\_02  
score 0.950  
sequence tcagtgatatggca

<221> protein\_bind  
<222> complement(478..489)  
<223> matinspector prediction  
name SRY\_02  
score 0.951  
sequence taaaacaaaaca



<221> protein\_bind  
 <222> 486..493  
 <223> matinspector prediction  
     name E2F\_02  
     score 0.957  
     sequence tttagcgc

<221> protein\_bind  
 <222> complement(514..521)  
 <223> matinspector prediction  
     name MZF1\_01  
     score 0.975  
     sequence tgagggga

<400> 17  
 tgagtgcagt gttacatgtc agttgggtta agtttggtta tgtcattcaa atcttctatg 60  
 tcttgatttg cctgctaatt ctattatttc tggaactaaa ttagtttgat ggttctatta 120  
 gttattgact gaggtgtgct aatctcccat tatgtggatt tatctatttc ttcagttgta 180  
 gataggacat tgatagatac ataagtacca ggacaaaagc agggagatct tttttccaaa 240  
 atcaggagaa aaaaatgaca tctggaaaac ctatagggaa aggcataaca gatggttaagg 300  
 atactttatc ttgagtagga gagccttcct gtggcaacgt ggagaaggga agaggtcgta 360  
 gaattgagga gtcagctcag ttagaagcag ggagttggga attccgttca tgtgatttag 420  
 catcagtgat atggcaaatg tgggactaag ggtagtgatc agaggggttaa aattgtgtgt 480  
 tttgttttag cgctgctggg gcacgcctt gggtcccctc aaacagattc ccatgaatct 540  
 cttcat 546

<210> 18  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <400> 18  
 gtaccaggga ctgtgacat tgc 23

<210> 19  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence  
 <400> 19  
 ctgtgacat tgctccaag agag 24

<210> 20  
 <211> 861  
 <212> DNA  
 <213> Homo Sapiens

<220>  
 <221> promoter  
 <222> 1..806  
 <221> transcription start site  
 <222> 807

<221> protein\_bind  
 <222> complement(60..70)  
 <223> matinspector prediction  
     name NFY\_Q6  
     score 0.956  
     sequence ggaccaatcat

<221> protein\_bind  
 <222> 70..77  
 <223> matinspector prediction  
     name MZF1\_01

```
score 0.962
sequence cctgggga

<221> protein_bind
<222> 124..132
<223> matinspector prediction
      name CMYB_01
      score 0.994
      sequence tgaccgttg

<221> protein_bind
<222> complement(126..134)
<223> matinspector prediction
      name VMYB_02
      score 0.985
      sequence tccaacggt

<221> protein_bind
<222> 135..143
<223> matinspector prediction
      name STAT_01
      score 0.968
      sequence ttcctggaa

<221> protein_bind
<222> complement(135..143)
<223> matinspector prediction
      name STAT_01
      score 0.951
      sequence ttccaggaa

<221> protein_bind
<222> complement(252..259)
<223> matinspector prediction
      name MZF1_01
      score 0.956
      sequence ttggggga

<221> protein_bind
<222> 357..368
<223> matinspector prediction
      name IK2_01
      score 0.965
      sequence gaatgggatttc

<221> protein_bind
<222> 384..391
<223> matinspector prediction
      name MZF1_01
      score 0.986
      sequence agagggga

<221> protein_bind
<222> complement(410..421)
<223> matinspector prediction
      name SRY_02
      score 0.955
      sequence gaaaacaaaaca

<221> protein_bind
<222> 592..599
<223> matinspector prediction
      name MZF1_01
```

score 0.960  
sequence gaagggga

<221> protein\_bind  
<222> 618..627  
<223> matinspector prediction  
name MYOD\_Q6  
score 0.981  
sequence agcatctgcc

<221> protein\_bind  
<222> 632..642  
<223> matinspector prediction  
name DELTAEF1\_01  
score 0.958  
sequence tcccaccttcc

<221> protein\_bind  
<222> complement(813..823)  
<223> matinspector prediction  
name S8\_01  
score 0.992  
sequence gaggaattat

<221> protein\_bind  
<222> complement(824..831)  
<223> matinspector prediction  
name MZF1\_01  
score 0.986  
sequence agagggga

<221> misc\_feature  
<222> 335,376  
<223> n=a, g, c or t

<400> 20  
tactataggg cacgcgtggt cgacggccgg gctgttctgg agcagagggc atgtcagtaa 60  
tgattggtcc ctggggaagg tctggctggc tccagcacag tgaggcattt aggtatctct 120  
cggtagaccgt tggattcctg gaagcagtag ctgttctggt tggatctggg agggacaggg 180  
ctcagagggc taggcacgag ggaagggtcag aggagaaggs aggsarggcc cagtgaarg 240  
gagcatgcc ttccccaac cctggcttsc ycttgymam agggcgkty tgggmacttr 300  
aaytcagggc ccaascagaa scacaggccc aktcntggct smaagcacia tagcctgaat 360  
gggatttcag gttagncagg gtgagagggg aggctctctg gcttagtttt gttttgtttt 420  
ccaaatcaag gtaacttgct cccttctgct acgggccttg gtcttggett gtcctcacc 480  
agtcggaact ccctaccact ttcaggagag tggttttagg cccgtggggc tgttctgttc 540  
caagcagtgt gagaacatgg ctggtagagg ctctagctgt gtgcggggcc tgaaggggag 600  
tgggttctcg cccaaagagc atctgcccac ttcccacctt cccttctccc accagaagct 660  
tgctgagct gtttgacaa aaatccaaac ccacttggc tactctggcc tggcttcagc 720  
ttggaacca atacctaggc ttacaggcca tctgagcca ggggcctctg gaaattctct 780  
tctgatggt cctttagggt tgggcacaaa atataattgc ctctccctc tccattttc 840  
tctcttggga gcaatggtca c 861

<210> 21  
<211> 20  
<212> DNA  
<213> Artificial Sequence  
<400> 21  
ctgggatgga aggcacggta

20

<210> 22  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<400> 22  
gagaccacac agctagacaa

20

<210> 23  
<211> 555  
<212> DNA  
<213> Homo Sapiens

<220>  
<221> promoter  
<222> 1..500

<221> transcription start site  
<222> 501

<221> protein\_bind  
<222> 191..206  
<223> matinspector prediction  
name ARNT\_01  
score 0.964  
sequence ggactcacgtgctgct

<221> protein\_bind  
<222> 193..204  
<223> matinspector prediction  
name NMYC\_01  
score 0.965  
sequence actcacgtgctg

<221> protein\_bind  
<222> 193..204  
<223> matinspector prediction  
name USF\_01  
score 0.985  
sequence actcacgtgctg

<221> protein\_bind  
<222> complement(193..204)  
<223> matinspector prediction  
name USF\_01  
score 0.985  
sequence cagcacgtgagt

<221> protein\_bind  
<222> complement(193..204)  
<223> matinspector prediction  
name NMYC\_01  
score 0.956  
sequence cagcacgtgagt

<221> protein\_bind  
<222> complement(193..204)  
<223> matinspector prediction  
name MYCMAX\_02  
score 0.972  
sequence cagcacgtgagt

<221> protein\_bind  
<222> 195..202  
<223> matinspector prediction  
name USF\_C  
score 0.997  
sequence tcacgtgc

<221> protein\_bind  
 <222> complement(195..202)  
 <223> matinspector prediction  
     name USF\_C  
     score 0.991  
     sequence gcacgtga

<221> protein\_bind  
 <222> complement(210..217)  
 <223> matinspector prediction  
     name MZF1\_01  
     score 0.968  
     sequence catgggga

<221> protein\_bind  
 <222> 397..410  
 <223> matinspector prediction  
     name ELK1\_02  
     score 0.963  
     sequence ctctccggaagcct

<221> protein\_bind  
 <222> 400..409  
 <223> matinspector prediction  
     name CETS1P54\_01  
     score 0.974  
     sequence tccggaagcc

<221> protein\_bind  
 <222> complement(460..470)  
 <223> matinspector prediction  
     name AP1\_Q4  
     score 0.963  
     sequence agtgactgaac

<221> protein\_bind  
 <222> complement(460..470)  
 <223> matinspector prediction  
     name AP1FJ\_Q2  
     score 0.961  
     sequence agtgactgaac

<221> protein\_bind  
 <222> 547..555  
 <223> matinspector prediction  
     name PADS\_C  
     score 1.000  
     sequence tgtggtctc

<400> 23  
 ctatagggca cgcktggtcg acggcccggg ctggtctggt ctgkgtgga gtcgggtga 60  
 aggacagcat ttgtkacatc tgggtctactg caccttccct ctgccgtgca cttggccttt 120  
 kawaagctca gcaccgggtgc ccatcacagg gccggcagca cacacatccc attactcaga 180  
 aggaactgac ggactcacgt gctgctccgt ccccatgagc tcagtggacc tgtctatgta 240  
 gagcagtcag acagtgcctg ggatagagt agagttcagc cagtaaatcc aagtgattgt 300  
 cattcctgtc tgcattagta actcccaacc tagatgtgaa aacttagttc tttctcatag 360  
 gttgctctgc ccatgggtccc actgcagacc caggcactct ccggaagcct ggaaatcacc 420  
 cgtgtcttct gctgtctccc gctcacatcc cacacttggt ttcagtcact gagttacaga 480  
 ttttgctctc tcaatttctc ttgtcttagt cccatcctct gttcccctgg ccagtttgtc 540  
 tagctgtgtg gtctc 555

<210> 24

<211> 251  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 13..249

<221> sig\_peptide  
 <222> 13..81  
 <223> Von Heijne matrix  
 score 11.8000001907349  
 seq CLFVCLFLSQSFA/FV

<400> 24  
 aaaagtattg gg atg cct agt tac aar gtg tgt ggg gtt ttt tgt ttg ttt 51  
 Met Pro Ser Tyr Lys Val Cys Gly Val Phe Cys Leu Phe  
 -20 -15  
 gtt tgt ttg ttt ttg agc cag agt ttt gct ttt gtc ctc cag gct gga 99  
 Val Cys Leu Phe Leu Ser Gln Ser Phe Ala Phe Val Leu Gln Ala Gly  
 -10 -5 1 5  
 gtg cag tgg cgc gat ctc tgc tca ctg caa cct cag ctt ccc agg ttc 147  
 Val Gln Trp Arg Asp Leu Cys Ser Leu Gln Pro Gln Leu Pro Arg Phe  
 10 15 20  
 ggg cca tcc tcc tgc ctc agc ctc cca agt ggc tgg gac tgc agg cgc 195  
 Gly Pro Ser Ser Cys Leu Ser Leu Pro Ser Gly Trp Asp Cys Arg Arg  
 25 30 35  
 cca cca cca cgc ctg gct aat tct tgt gtt ttc ggt gga gac ggg gtt 243  
 Pro Pro Pro Arg Leu Ala Asn Ser Cys Val Phe Gly Gly Asp Gly Val  
 40 45 50  
 tca ccg gg 251  
 Ser Pro  
 55

<210> 25  
 <211> 274  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 35..274  
 <221> sig\_peptide  
 <222> 35..82  
 <223> Von Heijne matrix  
 score 14.8000001907349  
 seq SLPLLLLLLGAWA/IP

<400> 25  
 acagactaca cttgctgaac tggctcctgg ggcc atg agg ctg tca ctg cca ctg 55  
 Met Arg Leu Ser Leu Pro Leu  
 -15 -10  
 ctg ctg ctg ctg ctg gga gcc tgg gcc atc cca ggg ggc ctc ggg gac 103  
 Leu Leu Leu Leu Leu Gly Ala Trp Ala Ile Pro Gly Gly Leu Gly Asp  
 -5 1 5  
 agg gcg cca ctc aca gcc aca gcc cca caa ctg gat gat gag gag atg 151  
 Arg Ala Pro Leu Thr Ala Thr Ala Pro Gln Leu Asp Asp Glu Glu Met  
 10 15 20  
 tac tca gcc cac atg ccc gct cac ctg cgc tgt gat gcc tgc aga gct 199  
 Tyr Ser Ala His Met Pro Ala His Leu Arg Cys Asp Ala Cys Arg Ala  
 25 30 35  
 gtg gct tac cag gtg agt cct tca cca ctg tca cct gcc ctg ctc aca 247

21

Val Ala Tyr Gln Val Ser Pro Ser Pro Leu Ser Pro Ala Leu Leu Thr  
 40 45 50 55  
 ccc ctt ctc aag cca gcc ccc acc ggg 274  
 Pro Leu Leu Lys Pro Ala Pro Thr Gly  
 60

<210> 26  
 <211> 230  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 29..229

<221> sig\_peptide  
 <222> 29..94  
 <223> Von Heijne matrix  
 score 13.8000001907349  
 seq LGLLLLWLRGARC/GV

<400> 26  
 aaggagtcag tctcagtcag gacacagc atg gac atg agg gtc ccc gct cag 52  
 Met Asp Met Arg Val Pro Ala Gln  
 -20 -15  
 ctc ctg ggg ctc ctg cta ctc tgg ctc cga ggt gcc aga tgt ggc gtc 100  
 Leu Leu Gly Leu Leu Leu Leu Trp Leu Arg Gly Ala Arg Cys Gly Val  
 -10 -5 1  
 cag atg acc cag ttt cca ctg tcc ctg tct gca tcg gta gga gac aga 148  
 Gln Met Thr Gln Phe Pro Leu Ser Leu Ser Ala Ser Val Gly Asp Arg  
 5 10 15  
 gtc acc atc act tgc cgg aca agc cat ata att aac atc ttt tta aat 196  
 Val Thr Ile Thr Cys Arg Thr Ser His Ile Ile Asn Ile Phe Leu Asn  
 20 25 30  
 tgg tat cag cag aaa cca ggc aaa gcc cct tgg g 230  
 Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Trp  
 35 40 45

<210> 27  
 <211> 195  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 44..193

<221> sig\_peptide  
 <222> 44..112  
 <223> Von Heijne matrix  
 score 13.8000001907349  
 seq VLLLLLLSGDVQS/SE

<400> 27  
 agaggggcttc cggggctgcc ggtctgagtg cagagctgct gtc atg gcg gcc gct 55  
 Met Ala Ala Ala  
 -20  
 ctg tgg ggc ttc ttt ccc gtc ctg ctg ctg ctg ctg cta tcg ggg gat 103  
 Leu Trp Gly Phe Phe Pro Val Leu Leu Leu Leu Leu Leu Ser Gly Asp  
 -15 -10 -5  
 gtc cag agc tcg gag gtg ccc ggg gct gct gct gag gga tcg gga ggg 151  
 Val Gln Ser Ser Glu Val Pro Gly Ala Ala Ala Glu Gly Ser Gly Gly  
 1 5 10

22

agt ggg gtc ggc ata gga gak cgc ttc aag att gag gga ctg gg 195  
 Ser Gly Val Gly Ile Gly Xaa Arg Phe Lys Ile Glu Gly Leu  
 15 20 25

<210> 28  
 <211> 276  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 25..276

<221> sig\_peptide  
 <222> 25..90  
 <223> Von Heijne matrix  
 score 13.5  
 seq LGLLLLWLXGARC/DI

<400> 28  
 agtcagtctc agacaggaca cagc atg gac atg agg gtc ccc gct cag ctc 51  
 Met Asp Met Arg Val Pro Ala Gln Leu  
 -20 -15  
 ctg ggg ctc ctg cta ctc tgg ctc yka ggt gcc aga tgt gac atc cag 99  
 Leu Gly Leu Leu Leu Leu Trp Leu Xaa Gly Ala Arg Cys Asp Ile Gln  
 -10 -5 1  
 atg aca cag tct cca gtc ctg cct gca tct gta gga gac aga gtc acc 147  
 Met Thr Gln Ser Pro Val Leu Pro Ala Ser Val Gly Asp Arg Val Thr  
 5 10 15  
 atc act tgc cgg gca agt cag agc att ggc agc tat tta aac tgg tat 195  
 Ile Thr Cys Arg Ala Ser Gln Ser Ile Gly Ser Tyr Leu Asn Trp Tyr  
 20 25 30 35  
 cag cat aaa cca ggg cat gcc cct cgc ctc ctg atc tat gct gca act 243  
 Gln His Lys Pro Gly His Ala Pro Arg Leu Leu Ile Tyr Ala Ala Thr  
 40 45 50  
 act ttg tcg agg ggc ggs ccg gcc aga ttc agt 276  
 Thr Leu Ser Arg Gly Gly Pro Ala Arg Phe Ser  
 55 60

<210> 29  
 <211> 240  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 25..240

<221> sig\_peptide  
 <222> 25..120  
 <223> Von Heijne matrix  
 score 13.5  
 seq LLLLLLLPPPGSC/AG

<400> 29  
 agggcgctgc gcggcgccagc gaaa atg gcg gct tcc agg tgg gcg cgc aag 51  
 Met Ala Ala Ser Arg Trp Ala Arg Lys  
 -30 -25  
 gcc gtg gtc ctg ctt tgt gcc tct gac ctg ctg ctg ctg cta ctg 99  
 Ala Val Val Leu Leu Cys Ala Ser Asp Leu Leu Leu Leu Leu Leu  
 -20 -15 -10  
 cta cca ccg cct ggg tcc tgc gcc ggc cga agg tcg ccy dgg acg ccc 147  
 Leu Pro Pro Pro Gly Ser Cys Ala Gly Arg Arg Ser Pro Xaa Thr Pro



23

```

      -5              1              5
gac gag tct acc cca cct ccc cgg aag aag aag aag gat att cgc gat      195
Asp Glu Ser Thr Pro Pro Pro Arg Lys Lys Lys Lys Asp Ile Arg Asp
10              15              20              25
tac aat gat gca gac atg gcg cgt ctt ctg gag caa ggg gag ggg      240
Tyr Asn Asp Ala Asp Met Ala Arg Leu Leu Glu Gln Gly Glu Gly
      30              35              40

```

<210> 30  
 <211> 461  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 80..460

<221> sig\_peptide  
 <222> 80..136  
 <223> Von Heijne matrix  
       score 13.5  
       seq WVLLLLALLEGVQC/DV

<221> misc\_feature  
 <222> 280..281,311..313  
 <223> n=a, g, c or t

```

<400> 30
agctctcaga gaggtgcctt agccctggat tccaaggcat ttccacttgg tgatcagcac      60
tgaacacaga ggactcacc atg gag ttg ggg ctg tgc tgg gtt ctc ctt tta      112
                        Met Glu Leu Gly Leu Cys Trp Val Leu Leu Leu
                        -15              -10
gct ctt tta gaa ggt gtc caa tgt gac gtg gaa tta gtg gag tct ggg      160
Ala Leu Leu Glu Gly Val Gln Cys Asp Val Glu Leu Val Glu Ser Gly
      -5              1              5
ggc ggc ttg gtg cag cct gga ggg tct ctg aga ctt tcc tgt gca gcc      208
Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala
      10              15              20
tct gga ttc aat ttt agc act tat gag atg cat tgg atc cgc cag gct      256
Ser Gly Phe Asn Phe Ser Thr Tyr Glu Met His Trp Ile Arg Gln Ala
      25              30              35              40
cca ggg aag ggg ccg gag tgg gtn nca tat gtc agt ggt gga ggt gga      304
Pro Gly Lys Gly Pro Glu Trp Val Xaa Tyr Val Ser Gly Gly Gly Gly
      45              50              55
acc agh nnn aac gcv sac tct gtg aag ggc cga ttc acc atc tcc aga      352
Thr Xaa Xaa Asn Ala Xaa Ser Val Lys Gly Arg Phe Thr Ile Ser Arg
      60              65              70
gac aat gcc aac agt ttt gtg tat cta caa atg gac agt ctg cga gtc      400
Asp Asn Ala Asn Ser Phe Val Tyr Leu Gln Met Asp Ser Leu Arg Val
      75              80              85
gag gac acc gct ctc tat tac tgt gcg aga rgg gat tac gac ttc tgg      448
Glu Asp Thr Ala Leu Tyr Tyr Cys Ala Arg Xaa Asp Tyr Asp Phe Trp
      90              95              100
agt ggt tat tat a
Ser Gly Tyr Tyr
105

```

<210> 31  
 <211> 112  
 <212> DNA  
 <213> Homo sapiens

<220>

```

<221> CDS
<222> 28..111

<221> sig_peptide
<222> 28..84
<223> Von Heijne matrix
      score 13.3999996185303
      seq LLLLLSHCTGSLS/QP

<400> 31
aactgtgcat gtcaggctgt gtccacc atg gcc tgg act cct ctt ctt ctc ttg      54
                                   Met Ala Trp Thr Pro Leu Leu Leu Leu
                                   -15
ctc ctc tct cac tgc aca ggt tcc ctc tcc cag cct gtg ctg act cag      102
Leu Leu Ser His Cys Thr Gly Ser Leu Ser Gln Pro Val Leu Thr Gln
-10                               -5                               1                               5
cca cgc ggg g                                                                112
Pro Arg Gly

<210> 32
<211> 445
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 80..445

<221> sig_peptide
<222> 80..136
<223> Von Heijne matrix
      score 12.8000001907349
      seq WVFLVALLRGVQC/QV

<221> misc_feature
<222> 2,7
<223> n=a, g, c or t

<400> 32
anctctngga gaggagccca gcactagaag tcggcggtgt ttccattcgg tgatcagcac      60
tgaacacaga ggactcacc atg gag ttt ggg ctg aat tgg gtt ttc ctc gtt      112
                                   Met Glu Phe Gly Leu Asn Trp Val Phe Leu Val
                                   -15                               -10
gct ctt tta aga ggt gtc cag tgt cag gtt cag ttg gtg gag tct ggg      160
Ala Leu Leu Arg Gly Val Gln Cys Gln Val Gln Leu Val Glu Ser Gly
-5                               1                               5
gga ggc gtg gtc cag cct ggg acg tcc ctg aca ctt tcc tgt gca ggc      208
Gly Gly Val Val Gln Pro Gly Thr Ser Leu Thr Leu Ser Cys Ala Gly
10                               15                               20
tcg gga ttc agt ttc agt gat tat ggc atc cac tgg gtc cgc cag gct      256
Ser Gly Phe Ser Phe Ser Asp Tyr Gly Ile His Trp Val Arg Gln Ala
25                               30                               35                               40
cca ggc aag ggg ctg gaa tgg gtg gcg gtt att tca cac gat gga aat      304
Pro Gly Lys Gly Leu Glu Trp Val Ala Val Ile Ser His Asp Gly Asn
45                               50                               55
aac aaa tat tat gga ggc tcc atg aag ggc cga gtc acc atc tcc aga      352
Asn Lys Tyr Tyr Gly Gly Ser Met Lys Gly Arg Val Thr Ile Ser Arg
60                               65                               70
gac aac tcc agg cat acc gtg tct ttg caa atg agc agc ttg gga cct      400
Asp Asn Ser Arg His Thr Val Ser Leu Gln Met Ser Ser Leu Gly Pro
75                               80                               85
gag gac acg gca gtg tat tac tgt gcg aaa gat cga acc ggg ggg      445
Glu Asp Thr Ala Val Tyr Tyr Cys Ala Lys Asp Arg Thr Gly Gly

```

90 95 100

<210> 33  
 <211> 321  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 41..319

<221> sig\_peptide  
 <222> 41..97  
 <223> Von Heijne matrix  
 score 12.6000003814697  
 seq FLLLLAAPRWVLS/QV

<400> 33  
 aaataacttts kgcagagtcc tggacctcct gtgcaagaac atg aaa ctt ctg tgg 55  
 Met Lys Leu Leu Trp  
 -15  
 ttc ttc ctt ctc ctg ctg gca gct ccc aga tgg gtc ctg tcc cag gtg 103  
 Phe Phe Leu Leu Leu Leu Ala Ala Pro Arg Trp Val Leu Ser Gln Val  
 -10 -5 1  
 cag ctg gtg smg tgc ggc cca gga ctg gtg aag cct tgc ggg acc ctg 151  
 Gln Leu Val Xaa Ser Gly Pro Gly Leu Val Lys Pro Ser Gly Thr Leu  
 5 10 15  
 tcc cta acg tgc act gts ksb ggk grs ksc ata act aat tac tac tgg 199  
 Ser Leu Thr Cys Thr Val Xaa Gly Xaa Xaa Ile Thr Asn Tyr Tyr Trp  
 20 25 30  
 agt bgg atc cgg cag tcc cca ggg aag gga ctg gag tgg att ggg act 247  
 Ser Xaa Ile Arg Gln Ser Pro Gly Lys Gly Leu Glu Trp Ile Gly Thr  
 35 40 45 50  
 atc tac tac agt ggg agc gcc gac cac aac ccc tcc ctc agg agt mga 295  
 Ile Tyr Tyr Ser Gly Ser Ala Asp His Asn Pro Ser Leu Arg Ser Arg  
 55 60 65  
 gcc act att tca tta gac acg cgc gg 321  
 Ala Thr Ile Ser Leu Asp Thr Arg  
 70

<210> 34  
 <211> 193  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 49..192

<221> sig\_peptide  
 <222> 49..108  
 <223> Von Heijne matrix  
 score 12.5  
 seq LLXLLTALPPLWS/SS

<400> 34  
 agagctcagg gtgckgagcg tgtgaccagc agtgagcaga ggccggcc atg gcc agc 57  
 Met Ala Ser  
 -20  
 ctg ggg ctg ctg ctc ctg ckc tta ctg aca gca ctg cca ccg ctg tgg 105  
 Leu Gly Leu Leu Leu Xaa Leu Leu Thr Ala Leu Pro Pro Leu Trp  
 -15 -10 -5  
 tcc tcc tca ctg cct ggg ctg gac ack gct gaa agt aaa gcc acc akt 153

26

Ser	Ser	Ser	Leu	Pro	Gly	Leu	Asp	Thr	Ala	Glu	Ser	Lys	Ala	Thr	Xaa	
1					5					10					15	
gca	gac	ctg	atc	ctg	tct	gcg	ctg	gag	aga	gcc	acc	ggg	g			193
Ala	Asp	Leu	Ile	Leu	Ser	Ala	Leu	Glu	Arg	Ala	Thr	Gly				
				20					25							

<210> 35  
 <211> 438  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 151..438

<221> sig\_peptide  
 <222> 151..234  
 <223> Von Heijne matrix  
 score 12.5  
 seq LLLLLLLPLRGQA/NT

<400> 35	
acgagaaggaggaggggggccc agccctgctt tgggcaatcc ttgctctgac cactcagaca	60
ccgtgtcctc ttgcctggga gaggggaagc agatctgagg acatctctgt gccaggccag	120
aaaccgcccc cctgcagttc cttctccggg atg gac gtg ggg ccc agc tcc ctg	174
Met Asp Val Gly Pro Ser Ser Leu	

ccc cac ctt ggg ctg aag ctg ctg ctg ctc ctg ctg ctg ctg ccc ctc	222
Pro His Leu Gly Leu Lys Leu Leu Leu Leu Leu Leu Leu Pro Leu	
-20 -15 -10 -5	
agg ggc caa gcc aac aca ggc tgc tac ggg atc cca ggg atg ccc ggc	270
Arg Gly Gln Ala Asn Thr Gly Cys Tyr Gly Ile Pro Gly Met Pro Gly	
1 5 10	
ctg ccc ggg gca cca ggg aag gat ggg tac gac gga ctg ccg ggg ccc	318
Leu Pro Gly Ala Pro Gly Lys Asp Gly Tyr Asp Gly Leu Pro Gly Pro	
15 20 25	
aag ggg gag cca gga atc cca gcc att ccc ggg atc cga gga ccc aaa	366
Lys Gly Glu Pro Gly Ile Pro Ala Ile Pro Gly Ile Arg Gly Pro Lys	
30 35 40	
ggg cag aag gga gaa ccc ggc tta ccc ggc cat cct ggg aaa aat ggc	414
Gly Gln Lys Gly Glu Pro Gly Leu Pro Gly His Pro Gly Lys Asn Gly	
45 50 55 60	
ccc atg gga ccc cct ggg atg cca	438
Pro Met Gly Pro Pro Gly Met Pro	
65	

<210> 36  
 <211> 488  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 59..487

<221> sig\_peptide  
 <222> 59..115  
 <223> Von Heijne matrix  
 score 12.3999996185303  
 seq ILLLVAAATGTHA/QV

<221> misc\_feature  
 <222> 26..28

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 36

```

atcacacaac agmcacatcs swmvsnnnmc agaagccccc agagtgcagc acctcacc      58
atg gac tgc acc tgg agg atc ctc ctc ttg gtg gca gca gct aca ggc      106
Met Asp Cys Thr Trp Arg Ile Leu Leu Val Ala Ala Ala Thr Gly
               -15               -10               -5
acc cac gcc cag gtc cag ttg gta cag tct ggg cct gag gtg aaa aag      154
Thr His Ala Gln Val Gln Leu Val Gln Ser Gly Pro Glu Val Lys Lys
               1               5               10
cct ggg gcc tca gtg aag gtc tcc tgc cag gtt tcc gga tac aac gtc      202
Pro Gly Ala Ser Val Lys Val Ser Cys Gln Val Ser Gly Tyr Asn Val
               15               20               25
gtg gaa tta tcc atc cac tgg gtg cgt cag tcg cct gga aaa ggg ctt      250
Val Glu Leu Ser Ile His Trp Val Arg Gln Ser Pro Gly Lys Gly Leu
               30               35               40               45
gag tgg atg gga ggt ttt gac ctt gaa agt ggt gaa aca atc tac gca      298
Glu Trp Met Gly Gly Phe Asp Leu Glu Ser Gly Glu Thr Ile Tyr Ala
               50               55               60
cag agg ttc cag ggc aga atc acc atg acc gag gac tca tct tca gac      346
Gln Arg Phe Gln Gly Arg Ile Thr Met Thr Glu Asp Ser Ser Ser Asp
               65               70               75
aca gcc ttc atg gag ctg atc agc ctg aga cct gaa gat gcg gcc gtc      394
Thr Ala Phe Met Glu Leu Ile Ser Leu Arg Pro Glu Asp Ala Ala Val
               80               85               90
tac tac tgt gca acg atc cgg ctg cca gta gtg ctt ttt ttc gcg gct      442
Tyr Tyr Cys Ala Thr Ile Arg Leu Pro Val Val Leu Phe Phe Ala Ala
               95               100               105
tct ggg gcc agg gaa ccc tgg tcg ccg tct cct cag cmt cca cgg g      488
Ser Gly Ala Arg Glu Pro Trp Ser Pro Ser Pro Gln Xaa Pro Arg
110               115               120

```

&lt;210&gt; 37

&lt;211&gt; 138

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 26..136

&lt;221&gt; sig\_peptide

&lt;222&gt; 26..79

&lt;223&gt; Von Heijne matrix

score 12.1000003814697

seq VLLLAVALLLAVLC/KV

&lt;400&gt; 37

```

ttttaccga cccgacgccg gcgtg atg tgg ctt ccg ctg gtg ctg ctc ctg      52
                               Met Trp Leu Pro Leu Val Leu Leu Leu
                               -15                               -10
gct gtg ctg ctg ctg gcc gtc ctc tgc aaa gtt tac ttg gga cta ttc      100
Ala Val Leu Leu Leu Ala Val Leu Cys Lys Val Tyr Leu Gly Leu Phe
                               -5                               1                               5
tct ggc agc tcc ccg aat cct ttc tcc gaa gaa agg gg      138
Ser Gly Ser Ser Pro Asn Pro Phe Ser Glu Glu Arg
10               15

```

&lt;210&gt; 38

&lt;211&gt; 163

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

<220>  
 <221> CDS  
 <222> 9..161

<221> sig\_peptide  
 <222> 9..83  
 <223> Von Heijne matrix  
 score 11.8999996185303  
 seq WLLLLPLLLGLNA/GA

<400> 38  
 aacttgatc atg gag ctg gca ctg cgg cgc tct ccc gtc ccg cgg tgg ttg 50  
 Met Glu Leu Ala Leu Arg Arg Ser Pro Val Pro Arg Trp Leu  
 -25 -20 -15  
 ctg ctg ctg ccg ctg ctg ctg ggc ctg aac gca gga gct gtc att gac 98  
 Leu Leu Leu Pro Leu Leu Leu Gly Leu Asn Ala Gly Ala Val Ile Asp  
 -10 -5 1 5  
 tgg ccc aca gag gag ggc aag gaa gta tgg gat tat gtg acg gtc cgc 146  
 Trp Pro Thr Glu Glu Gly Lys Glu Val Trp Asp Tyr Val Thr Val Arg  
 10 15 20  
 aag gat gcc tac atg gg 163  
 Lys Asp Ala Tyr Met  
 25

<210> 39  
 <211> 427  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 35..427

<221> sig\_peptide  
 <222> 35..91  
 <223> Von Heijne matrix  
 score 11.8999996185303  
 seq FLFLLTCCPGSNS/QA

<221> misc\_feature  
 <222> 138..139  
 <223> n=a, g, c or t

<400> 39  
 tctggcacca ggggtccctt ccaatatcag cacc atg gcc tgg act cct ctc ttt 55  
 Met Ala Trp Thr Pro Leu Phe  
 -15  
 ctg ttc ctc ctc act tgc tgc cca ggg tcc aat tcc cag gct gtg gkg 103  
 Leu Phe Leu Leu Thr Cys Cys Pro Gly Ser Asn Ser Gln Ala Val Xaa  
 -10 -5 1  
 act cag gag ccc ctc act gac tgt gtc ccc cgg ann aca gtc act ctc 151  
 Thr Gln Glu Pro Leu Thr Asp Cys Val Pro Arg Xaa Thr Val Thr Leu  
 5 10 15 20  
 acc tgt ggc tcc agt att gga gct gtc acc aat ggt cat ttt ccc tac 199  
 Thr Cys Gly Ser Ser Ile Gly Ala Val Thr Asn Gly His Phe Pro Tyr  
 25 30 35  
 tgg ttc caa cag aag cct ggc caa gcc ccc agg aca ctg att tct gat 247  
 Trp Phe Gln Gln Lys Pro Gly Gln Ala Pro Arg Thr Leu Ile Ser Asp  
 40 45 50  
 acg ttc aac aga cag tcc tcg aca cct gcc cgc ttc tct ggc tcc ctc 295  
 Thr Phe Asn Arg Gln Ser Ser Thr Pro Ala Arg Phe Ser Gly Ser Leu  
 55 60 65  
 ctg ggg ggc aaa gct gtc ctg act ctt tcg gat gcg caa cct gac gat 343

29

Leu	Gly	Gly	Lys	Ala	Val	Leu	Thr	Leu	Ser	Asp	Ala	Gln	Pro	Asp	Asp		
	70					75					80						
gag	gct	gaa	tat	tat	tgt	gtc	ctc	tcc	tat	agt	ggg	ggg	cgg	ccg	gtg	391	
Glu	Ala	Glu	Tyr	Tyr	Cys	Val	Leu	Ser	Tyr	Ser	Gly	Gly	Arg	Pro	Val		
85					90					95					100		
ttc	ggc	gga	ggg	acc	aag	ctg	acc	gtc	cta	agt	cag					427	
Phe	Gly	Gly	Gly	Thr	Lys	Leu	Thr	Val	Leu	Ser	Gln						
				105					110								

<210> 40  
 <211> 97  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 22..96

<221> sig\_peptide  
 <222> 22..84  
 <223> Von Heijne matrix  
 score 11.8999996185303  
 seq LALCLLLGPLAGA/KP

<400> 40																	
agatcaggaa	gcaccgggaa	g	atg	cag	gcc	tgc	atg	gtg	ccg	ggg	ctg	gcc				51	
			Met	Gln	Ala	Cys	Met	Val	Pro	Gly	Leu	Ala					
			-20						-15								
ctc	tgc	ctc	cta	ctg	ggg	cct	ctt	gca	ggg	gcc	aag	cct	gtg	cag	g	97	
Leu	Cys	Leu	Leu	Leu	Gly	Pro	Leu	Ala	Gly	Ala	Lys	Pro	Val	Gln			
-10						-5					1						

<210> 41  
 <211> 536  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 266..535

<221> sig\_peptide  
 <222> 266..307  
 <223> Von Heijne matrix  
 score 15  
 seq LLPLLLLLLPMCWA/VE

<400> 41																	
acttttgagg	tcacgtgctc	attccgtttc	cctacctccc	ccaaccttat	cccgcccctg											60	
gggggttcg	ggcatttttc	aggaactttc	tttccggctt	gagaagccgc	cactcccaag											120	
atgsagcagg	aaccgcggct	gctggacaag	aggggtgcgg	tggatactga	cctttgctcc											180	
ggcctcgtcg	tgaagacaca	gcgcactctc	ccgctgtagg	cttcctccca	cagaaccctg											240	
ttcgggcctc	agagcgtctg	gtgag	atg	ctg	ttg	ccg	ctg	ctg	ctg	cta						292	
			Met	Leu	Leu	Pro	Leu	Leu	Leu	Leu							
			-10														
ccc	atg	tgc	tgg	gcc	gtg	gag	gtc	aag	agg	ccc	cgg	ggc	gtc	tcc	ctc	340	
Pro	Met	Cys	Trp	Ala	Val	Glu	Val	Lys	Arg	Pro	Arg	Gly	Val	Ser	Leu		
-5				1				5					10				
acc	aat	cat	cac	ttc	tac	gat	gag	tcc	aag	cct	ttc	acc	tgc	ctg	gac	388	
Thr	Asn	His	His	Phe	Tyr	Asp	Glu	Ser	Lys	Pro	Phe	Thr	Cys	Leu	Asp		
			15				20				25						
ggg	tcg	gcc	acc	atc	cca	ttt	gat	cag	gtc	aac	gat	gac	tat	tgc	gac	436	
Gly	Ser	Ala	Thr	Ile	Pro	Phe	Asp	Gln	Val	Asn	Asp	Asp	Tyr	Cys	Asp		

30

30	35	40	
tgc aaa gat ggc tct gac gag cca ggc acg gct gcc tgt cct aat ggc	484		
Cys Lys Asp Gly Ser Asp Glu Pro Gly Thr Ala Ala Cys Pro Asn Gly			
45	50	55	
agc ttc cac tgc acc aac act ggc tat aag ccc ctg tat atc ccc tcc	532		
Ser Phe His Cys Thr Asn Thr Gly Tyr Lys Pro Leu Tyr Ile Pro Ser			
60	65	70	75
aac c			
Asn			536

<210> 42  
 <211> 319  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 143..319

<221> sig\_peptide  
 <222> 143..205  
 <223> Von Heijne matrix  
 score 11.6000003814697  
 seq LLLCLALSGAAET/KP

<221> misc\_feature  
 <222> 139  
 <223> n=a, g, c or t

<400> 42	
agcagagggga acaggggaaga aacctaagg ctgcaggtctg ccaggtgtgc ttggagagcc	60
cccttcttcc gccgggcctc gcaagcagcg taggactgtg gagaagggcg gtgggcaagg	120
agggaaactcg agagcarny cc atg ggc aca cag gag ggc tgg wgc ctg ctg	172
Met Gly Thr Gln Glu Gly Trp Xaa Leu Leu	
-20	-15
ctc tgc ctg gct cta tct gga gca gca gaa acc aag ccc cac cca gca	220
Leu Cys Leu Ala Leu Ser Gly Ala Ala Glu Thr Lys Pro His Pro Ala	
-10	-5
gag ggg cag tgg cgg gca gtg gdc gtg gtc cta gac ygt ttc ctg gtg	268
Glu Gly Gln Trp Arg Ala Val Xaa Val Val Leu Asp Xaa Phe Leu Val	
10	15
aag gac svt gcg cac cgt gga gct ctc gcc agc agt gag gac agg gca	316
Lys Asp Xaa Ala His Arg Gly Ala Leu Ala Ser Ser Glu Asp Arg Ala	
25	30
agg	319
Arg	

<210> 43  
 <211> 412  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 35..412

<221> sig\_peptide  
 <222> 35..82  
 <223> Von Heijne matrix  
 score 11.1999998092651  
 seq LVVFLLWGVTVG/PV

<221> misc\_feature



&lt;222&gt; 148

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 43

```

agacactcac tgcaccggag tgagcgcgac catc atg tcc atg ctc gtg gtc ttt      55
                                   Met Ser Met Leu Val Val Phe
                                   -15                -10
ctc ttg ctg tgg ggt gtc acc tgg ggc cca gtg aca gaa gca gcc ata      103
Leu Leu Leu Trp Gly Val Thr Trp Gly Pro Val Thr Glu Ala Ala Ile
                                   -5                1                5
ttt tat gag acg cag scc agc ctg tgg gca gag tcc gaa cac tgn ctg      151
Phe Tyr Glu Thr Gln Xaa Ser Leu Trp Ala Glu Ser Glu His Xaa Leu
                                   10                15                20
aaa acc ctt ggc caa tgt gac gct gac gtg cca ggc ccg cct gga gac      199
Lys Thr Leu Gly Gln Cys Asp Ala Asp Val Pro Gly Pro Pro Gly Asp
                                   25                30                35
tcc aga ctt cca gct gtt caa gaa tgg ggg gcc cag gag cct gtg cac      247
Ser Arg Leu Pro Ala Val Gln Glu Trp Gly Ala Gln Glu Pro Val His
                                   40                45                50                55
ctt gac tca cct gcc atc aag cac cag ttc ctg ctg acg ggt gac acc      295
Leu Asp Ser Pro Ala Ile Lys His Gln Phe Leu Leu Thr Gly Asp Thr
                                   60                65                70
cag ggc cgc tac cgc tgc cgc tgc ggc ttg tcc aca gga tgg mcc cag      343
Gln Gly Arg Tyr Arg Cys Arg Ser Gly Leu Ser Thr Gly Trp Xaa Gln
                                   75                80                85
ctg agc aag ctc ctg gag ctg aca ggg cca aaa gtc ctt gcc tgc tcc      391
Leu Ser Lys Leu Leu Glu Leu Thr Gly Pro Lys Val Leu Ala Cys Ser
                                   90                95                100
ctg gct ctc gat ggc gcc agt      412
Leu Ala Leu Asp Gly Ala Ser
                                   105                110

```

&lt;210&gt; 44

&lt;211&gt; 331

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 32..331

&lt;221&gt; sig\_peptide

&lt;222&gt; 32..88

&lt;223&gt; Von Heijne matrix

score 11.1999998092651

seq IGFLLLWVPASRG/EI

&lt;400&gt; 44

```

atgagcaaaa ctgacaagtc aaggcaggaa g atg ttg cca tca caa ctc att      52
                                   Met Leu Pro Ser Gln Leu Ile
                                   -15
ggg ttt ctg ctg ctc tgg gtt cca gcc tcc agg ggt gaa att gtg ctg      100
Gly Phe Leu Leu Trp Val Pro Ala Ser Arg Gly Glu Ile Val Leu
                                   -10                -5                1
act cag tct cca gac ttt ctg tct gtg act cca aag gag aaa gtc acc      148
Thr Gln Ser Pro Asp Phe Leu Ser Val Thr Pro Lys Glu Lys Val Thr
                                   5                10                15                20
atc acc tgc cgg gcc agt sag agc att ggt agt agt tta tac tgg tac      196
Ile Thr Cys Arg Ala Ser Xaa Ser Ile Gly Ser Ser Leu Tyr Trp Tyr
                                   25                30                35
cag cag aaa cca cat cag tct cca aag ctc gtc atc aag tat gct tcc      244
Gln Gln Lys Pro His Gln Ser Pro Lys Leu Val Ile Lys Tyr Ala Ser
                                   40                45                50

```

32

cag	tcc	ttc	tca	ggg	gtc	tcc	tcg	agg	ttc	agt	ggc	agt	gga	tct	ggg	292
Gln	Ser	Phe	Ser	Gly	Val	Ser	Ser	Arg	Phe	Ser	Gly	Ser	Gly	Ser	Gly	
		55					60					65				
aca	gat	ttc	acc	ctc	aca	atc	aat	agc	ctg	gaa	cct	ggg				331
Thr	Asp	Phe	Thr	Leu	Thr	Ile	Asn	Ser	Leu	Glu	Pro	Gly				
	70					75					80					

&lt;210&gt; 45

&lt;211&gt; 520

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 176..520

&lt;221&gt; sig\_peptide

&lt;222&gt; 176..235

&lt;223&gt; Von Heijne matrix

score 11.1999998092651

seq AFLLLVALSYTLA/RD

&lt;400&gt; 45

gaagataatc	acttgggggaa	aggaagggttc	gtttctgagt	tagcaacaag	taaattgcagc	60
actagtgggt	gggattgagg	tatgccctgg	tgcatataaata	gagactcagc	tgtgctggca	120
cactcagaag	cttggaaccgc	atcctagccg	ccgactcaca	caaggcagag	ttgcc atg	178
					Met	
					-20	

gag	aaa	att	cca	gtg	tca	gca	ttc	ttg	ctc	ctt	gtg	gcc	ctc	tcc	tac	226
Glu	Lys	Ile	Pro	Val	Ser	Ala	Phe	Leu	Leu	Leu	Val	Ala	Leu	Ser	Tyr	
			-15					-10					-5			

act	ctg	gcc	aga	gat	acc	aca	gtc	aaa	cct	gga	gcc	aaa	aag	gac	aca	274
Thr	Leu	Ala	Arg	Asp	Thr	Thr	Val	Lys	Pro	Gly	Ala	Lys	Lys	Asp	Thr	
		1				5					10					

aag	gac	tct	cga	ccc	aaa	ctg	ccc	cag	acc	ctc	tcc	aga	ggg	tgg	ggg	322
Lys	Asp	Ser	Arg	Pro	Lys	Leu	Pro	Gln	Thr	Leu	Ser	Arg	Gly	Trp	Gly	
	15				20					25						

gac	caa	ctc	atc	tgg	act	cag	aca	tat	gaa	gaa	gct	cta	tat	aaa	tcc	370
Asp	Gln	Leu	Ile	Trp	Thr	Gln	Thr	Tyr	Glu	Glu	Ala	Leu	Tyr	Lys	Ser	
	30				35				40					45		

aag	aca	agc	aac	aaa	ccc	ttg	atg	att	att	cat	cac	ttg	gat	gag	tgc	418
Lys	Thr	Ser	Asn	Lys	Pro	Leu	Met	Ile	Ile	His	His	Leu	Asp	Glu	Cys	
			50					55					60			

cca	cac	agt	caa	gct	tta	aag	aaa	gtg	ttt	gct	gaa	aat	aaa	gaa	atc	466
Pro	His	Ser	Gln	Ala	Leu	Lys	Lys	Val	Phe	Ala	Glu	Asn	Lys	Glu	Ile	
			65					70				75				

cag	aaa	ttg	gca	gag	cag	ttt	gtc	ctc	ctc	aat	ctg	gtt	tat	gaa	aca	514
Gln	Lys	Leu	Ala	Glu	Gln	Phe	Val	Leu	Leu	Asn	Leu	Val	Tyr	Glu	Thr	
		80					85					90				

act	gac															520
Thr	Asp															
	95															

&lt;210&gt; 46

&lt;211&gt; 383

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 25..381

&lt;221&gt; sig\_peptide

&lt;222&gt; 25..84

&lt;223&gt; Von Heijne matrix

score 11.1000003814697

seq LLALLFFLGQAAG/DL

&lt;400&gt; 46

```

agcggctcca gctaagagga caag atg agg ccc ggc ctc tca ttt ctc cta      51
                        Met Arg Pro Gly Leu Ser Phe Leu Leu
                        -20                      -15
gcc ctt ctg ttc ttc ctt ggc caa gct gca ggg gat ttg ggg gat gtg      99
Ala Leu Leu Phe Phe Leu Gly Gln Ala Ala Gly Asp Leu Gly Asp Val
-10                      -5                      1                      5
gga cct cca att ccc agc ccc ggc ttc agc tct ttc cca ggt gtt gac      147
Gly Pro Pro Ile Pro Ser Pro Gly Phe Ser Ser Phe Pro Gly Val Asp
                        10                      15                      20
tcc agc tcc agc ttc agc tcc agc tcc agg tcg ggc tcc agc tcc agc      195
Ser Ser Ser Ser Phe Ser Ser Ser Ser Arg Ser Gly Ser Ser Ser Ser
                        25                      30                      35
cgc agc tta ggc agc gga ggt tct gtg tcc cag ttg ttt tcc aat ttc      243
Arg Ser Leu Gly Ser Gly Gly Ser Val Ser Gln Leu Phe Ser Asn Phe
40                      45                      50
acc ggc tcc gtg gat gac cgt ggg acc tgc cag tgc tct gtt tcc ctg      291
Thr Gly Ser Val Asp Asp Arg Gly Thr Cys Gln Cys Ser Val Ser Leu
55                      60                      65
cca gac acc acc ttt ccc gtg gac aga gtg gaa cgc ttg gaa ttc aca      339
Pro Asp Thr Thr Phe Pro Val Asp Arg Val Glu Arg Leu Glu Phe Thr
70                      75                      80                      85
gct cat gtt ctt tct cag aag ttt gag aaa gaa ctt tct aaa gc      383
Ala His Val Leu Ser Gln Lys Phe Glu Lys Glu Leu Ser Lys
90                      95

```

&lt;210&gt; 47

&lt;211&gt; 459

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 17..457

&lt;221&gt; sig\_peptide

&lt;222&gt; 17..94

&lt;223&gt; Von Heijne matrix

score 11.1000003814697

seq FLLLVAAPRWVRS/QV

&lt;221&gt; misc\_feature

&lt;222&gt; 399

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 47

```

atactttctg agactc atg gac ctc ctg cac aag aac atg aaa cac ctg tgg      52
                        Met Asp Leu Leu His Lys Asn Met Lys His Leu Trp
                        -25                      -20                      -15
ttc ttc ctc ctc ctg gtg gca gct ccc aga tgg gtc cgg tct car gtg      100
Phe Phe Leu Leu Leu Val Ala Ala Pro Arg Trp Val Arg Ser Gln Val
-10                      -5                      1
cag ctg cak gag tcg ggc cca gga ctg gtg aag cct tcg ggg acc ctg      148
Gln Leu Xaa Glu Ser Gly Pro Gly Leu Val Lys Pro Ser Gly Thr Leu
5                      10                      15
tcc ctc atc tgc ggt gtc tct ggt gat tcc gtc acc att agt ggt tgg      196
Ser Leu Ile Cys Gly Val Ser Gly Asp Ser Val Thr Ile Ser Gly Trp
20                      25                      30

```

```

tgg agt tgg gtc cgc cag ccc cca ggg aag gga ctg gag tgg att tcg      244
Trp Ser Trp Val Arg Gln Pro Pro Gly Lys Gly Leu Glu Trp Ile Ser
35                               40                               45                               50
gaa atc gat cat ggt gga aac acc aat tac aac ccg tcc ctc aag agt      292
Glu Ile Asp His Gly Gly Asn Thr Asn Tyr Asn Pro Ser Leu Lys Ser
55                               60                               65
cga gtc kcc att tct tta gac aag tcc aag aat aag ttc tcc ctg agg      340
Arg Val Xaa Ile Ser Leu Asp Lys Ser Lys Asn Lys Phe Ser Leu Arg
70                               75                               80
ctg acc tct gtg acc gcc gcg gac acc gcc atg tat kac tgt gcg aga      388
Leu Thr Ser Val Thr Ala Ala Asp Thr Ala Met Tyr Xaa Cys Ala Arg
85                               90                               95
ggc ggt gcg bnc agc tcg tcc gct ttt gat gtc tgg ggc cta rgg aca      436
Gly Gly Ala Xaa Ser Ser Ser Ala Phe Asp Val Trp Gly Leu Xaa Thr
100                               105                               110
atg gtc atc atc tct tca gcc tc      459
Met Val Ile Ile Ser Ser Ala
115                               120

```

&lt;210&gt; 48

&lt;211&gt; 437

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 20..436

&lt;221&gt; sig\_peptide

&lt;222&gt; 20..76

&lt;223&gt; Von Heijne matrix

score 11

seq TLLLLTVPSWVLS/QV

&lt;400&gt; 48

```

gtgaatcctg ctctccacc atg gac ata ctt tgt tcc acg ctc ctg ctm ctg      52
Met Asp Ile Leu Cys Ser Thr Leu Leu Leu Leu
-15                               -10
ack gtc ccg tcc tgg gtc tta tcc car gtc acc ttg arg gaa tct ggt      100
Thr Val Pro Ser Trp Val Leu Ser Gln Val Thr Leu Xaa Glu Ser Gly
-5                               1                               5
cct gcg ctg gtg aaa gcc aca cag acc ctc aga ctg acc tgc acc ttc      148
Pro Ala Leu Val Lys Ala Thr Gln Thr Leu Arg Leu Thr Cys Thr Phe
10                               15                               20
tct ggg ttc tca ctc agc act aat aga atg cgt gtg agt tgg atc cgt      196
Ser Gly Phe Ser Leu Ser Thr Asn Arg Met Arg Val Ser Trp Ile Arg
25                               30                               35                               40
cag ccc cca ggg aag gcc ctg gag tgg ctt gca cgg att gat tgg gat      244
Gln Pro Pro Gly Lys Ala Leu Glu Trp Leu Ala Arg Ile Asp Trp Asp
45                               50                               55
gat tat aag agg tac agc aca tct ctg aag acc agg gtc acc atc tcc      292
Asp Tyr Lys Arg Tyr Ser Thr Ser Leu Lys Thr Arg Val Thr Ile Ser
60                               65                               70
aag gac acg tcc aaa aac cag gtg atc ctg aca atg acc aac gtg gac      340
Lys Asp Thr Ser Lys Asn Gln Val Ile Leu Thr Met Thr Asn Val Asp
75                               80                               85
cct gcg gac aca gcc acc tat tac tgt gca cgc ctt tca acg gca gct      388
Pro Ala Asp Thr Ala Thr Tyr Tyr Cys Ala Arg Leu Ser Thr Ala Ala
90                               95                               100
acc cca cag ttt ttt gac ttc tgg ggc cag gga gtc ctg gtc tcc gtc t      437
Thr Pro Gln Phe Phe Asp Phe Trp Gly Gln Gly Val Leu Val Ser Val
105                               110                               115                               120

```

<210> 49  
 <211> 456  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 40..456

<221> sig\_peptide  
 <222> 40..96  
 <223> Von Heijne matrix  
 score 10.8999996185303  
 seq FLLLVAAPRWVLS/QV

<400> 49  
 aaatactttc tgagagtcct ggacctcctg tgcaagaac atg adw cat ctg tgg 54  
 Met Xaa His Leu Trp  
 -15  
 ttc ttc ctt ctc ctg gtg gca gct ccc aga tgg gtc ctg tcc cag gtg 102  
 Phe Phe Leu Leu Leu Val Ala Ala Pro Arg Trp Val Leu Ser Gln Val  
 -10 -5 1  
 cag ctg cag gag tcg ggc cca gga ctg gtg aag cct tcg kwg acc ctg 150  
 Gln Leu Gln Glu Ser Gly Pro Gly Leu Val Lys Pro Ser Xaa Thr Leu  
 5 10 15  
 tcc ctc acc tgc act gtc tct ggt gac tcc atc agt agt tac tac tgg 198  
 Ser Leu Thr Cys Thr Val Ser Gly Asp Ser Ile Ser Ser Tyr Tyr Trp  
 20 25 30  
 agc tgg atc cgg cag ccc cca ggg aag gga ctg gag tgg att ggc tat 246  
 Ser Trp Ile Arg Gln Pro Pro Gly Lys Gly Leu Glu Trp Ile Gly Tyr  
 35 40 45 50  
 atc tat tac agt ggg agc acc aac tac aac ccc tcc ctc aag agt cga 294  
 Ile Tyr Tyr Ser Gly Ser Thr Asn Tyr Asn Pro Ser Leu Lys Ser Arg  
 55 60 65  
 gtc acc ata tca gtg gac acg tcc aag aac caa ttc tcc ctg aag ctg 342  
 Val Thr Ile Ser Val Asp Thr Ser Lys Asn Gln Phe Ser Leu Lys Leu  
 70 75 80  
 agc tct gtg acc gca gcg gac acg gcc gtg tat tac tgt gcg aga sgg 390  
 Ser Ser Val Thr Ala Ala Asp Thr Ala Val Tyr Tyr Cys Ala Arg Xaa  
 85 90 95  
 ctg cma tac tat gat agg agt ggt tat ttc aga tat ttt gac tac tgg 438  
 Leu Xaa Tyr Tyr Asp Arg Ser Gly Tyr Phe Arg Tyr Phe Asp Tyr Trp  
 100 105 110  
 ggc cag gga acc tgg tca 456  
 Gly Gln Gly Thr Trp Ser  
 115 120

<210> 50  
 <211> 447  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 38..445

<221> sig\_peptide  
 <222> 38..94  
 <223> Von Heijne matrix  
 score 10.8999996185303  
 seq FLLLVAAPRWVLS/QV

<221> misc\_feature

&lt;222&gt; 16

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 50

```

atactttctg agagtnctgg acctcctgtg caagaac atg aaa cat ctg tgg ttc      55
                                   Met Lys His Leu Trp Phe
                                   -15
ttc ctc ctc ctg gtg gca gct ccc aga tgg gtc ctg tcc cag gtg cag      103
Phe Leu Leu Leu Val Ala Ala Pro Arg Trp Val Leu Ser Gln Val Gln
                                   -10      -5      1
ctg cag gag tcg ggc cca gga ctg gtg aag cct tca cag acc ctg tcc      151
Leu Gln Glu Ser Gly Pro Gly Leu Val Lys Pro Ser Gln Thr Leu Ser
                                   5      10      15
ctc acc tgc aca gtc tct ggt ggc tcc atc gac agt ggt aat tac tac      199
Leu Thr Cys Thr Val Ser Gly Gly Ser Ile Asp Ser Gly Asn Tyr Tyr
20      25      30      35
tgg agc tgg atc cgg cag ccc gcc ggg aag gga ctg gag tgg att ggg      247
Trp Ser Trp Ile Arg Gln Pro Ala Gly Lys Gly Leu Glu Trp Ile Gly
                                   40      45      50
cgc atc tat agt act ggg agc acc aat tac aac ccc tcc ctc agc agt      295
Arg Ile Tyr Ser Thr Gly Ser Thr Asn Tyr Asn Pro Ser Leu Ser Ser
                                   55      60      65
cga gtc cag ata tcg tta gac acg tcc aag aac ctg ctc tcc ttg aac      343
Arg Val Gln Ile Ser Leu Asp Thr Ser Lys Asn Leu Leu Ser Leu Asn
                                   70      75      80
ctg acc tct gtg acc gcc gca gac acg gcc gtc tat ttt tgt gcg cga      391
Leu Thr Ser Val Thr Ala Ala Asp Thr Ala Val Tyr Phe Cys Ala Arg
                                   85      90      95
acc ttc ccc ttc tac tgg tac ctc gat ctc tgg ggc cgt ggc atc ctg      439
Thr Phe Pro Phe Tyr Trp Tyr Leu Asp Leu Trp Gly Arg Gly Ile Leu
100      105      110      115
gtc act gt      447
Val Thr

```

&lt;210&gt; 51

&lt;211&gt; 466

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 38..466

&lt;221&gt; sig\_peptide

&lt;222&gt; 38..94

&lt;223&gt; Von Heijne matrix

score 10.8999996185303

seq FLLLVAAPRWVLS/QV

&lt;221&gt; misc\_feature

&lt;222&gt; 423

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 51

```

atactttctg agagtcctgg acctcctgtg caagaac atg aaa cac ctg tgg ttc      55
                                   Met Lys His Leu Trp Phe
                                   -15
ttc ctc ctg ctg gtg gca gct ccc aga tgg gtc ctg tcc cag gtg cag      103
Phe Leu Leu Leu Val Ala Ala Pro Arg Trp Val Leu Ser Gln Val Gln
                                   -10      -5      1
ctg cag gag tcg ggc cca aga ctg gtg aag cct tca cag acc ctg tcc      151

```

37

Leu	Gln	Glu	Ser	Gly	Pro	Arg	Leu	Val	Lys	Pro	Ser	Gln	Thr	Leu	Ser		
5						10				15							
ctc	acc	tgc	act	gtc	tct	ggg	ggc	tcc	atc	agc	agt	ggg	ggg	tac	ttc		199
Leu	Thr	Cys	Thr	Val	Ser	Gly	Gly	Ser	Ile	Ser	Ser	Gly	Gly	Tyr	Phe		
20					25				30					35			
tgg	agt	tgg	atc	cgc	cag	cac	cca	ggg	cgg	ggc	ctg	gag	tgg	att	ggc		247
Trp	Ser	Trp	Ile	Arg	Gln	His	Pro	Gly	Arg	Gly	Leu	Glu	Trp	Ile	Gly		
				40				45						50			
tac	atc	tat	tac	aat	tgg	agc	acc	tac	tac	aat	ccg	tcc	ctc	agg	agt		295
Tyr	Ile	Tyr	Tyr	Asn	Trp	Ser	Thr	Tyr	Tyr	Asn	Pro	Ser	Leu	Arg	Ser		
				55				60						65			
cga	gtt	acc	atg	tca	atg	gac	acg	tct	aag	aac	cag	ttc	tcc	ctg	aac		343
Arg	Val	Thr	Met	Ser	Met	Asp	Thr	Ser	Lys	Asn	Gln	Phe	Ser	Leu	Asn		
				70				75						80			
ctg	aac	tct	gta	act	gcc	gcg	gac	acg	gsc	atg	tat	tac	tgt	gcs	aga		391
Leu	Asn	Ser	Val	Thr	Ala	Ala	Asp	Thr	Xaa	Met	Tyr	Tyr	Cys	Ala	Arg		
				85				90						95			
ggg	cgc	gga	cgc	ctt	ggc	tgg	ttc	ash	mct	tng	ggg	mca	ggg	rac	cca		439
Gly	Arg	Gly	Arg	Leu	Gly	Trp	Phe	Xaa	Xaa	Xaa	Gly	Xaa	Gly	Xaa	Pro		
100					105					110				115			
ggg	cac	cgt	ctc	atc	agc	cgt	cca	ggg									466
Gly	His	Arg	Leu	Ile	Ser	Arg	Pro	Gly									
				120													

&lt;210&gt; 52

&lt;211&gt; 392

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 59..391

&lt;221&gt; sig\_peptide

&lt;222&gt; 59..115

&lt;223&gt; Von Heijne matrix

score 10.8999996185303

seq FLLLVAAPRWVLS/QV

&lt;221&gt; misc\_feature

&lt;222&gt; 342

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 52

agggtcctgc	tcacatggga	aatactttct	gagagtcctg	gacctcctgt	gcaagaac												58
atg	aaa	cac	ctg	tgg	ttc	ttc	ctc	ctg	gtg	gca	gct	ccc	aga	tgg			106
Met	Lys	His	Leu	Trp	Phe	Phe	Leu	Leu	Leu	Val	Ala	Ala	Pro	Arg	Trp		
				-15				-10					-5				
gtc	ctg	tcc	cag	gtg	cag	ctg	cag	gag	tcg	ggc	cca	gga	ctg	gtg	aag		154
Val	Leu	Ser	Gln	Val	Gln	Leu	Gln	Glu	Ser	Gly	Pro	Gly	Leu	Val	Lys		
				1			5						10				
cct	tca	gag	acc	ctg	tcc	ctc	acc	tgc	act	gtc	tct	ggg	ggc	tcc	atc		202
Pro	Ser	Glu	Thr	Leu	Ser	Leu	Thr	Cys	Thr	Val	Ser	Gly	Gly	Ser	Ile		
				15			20						25				
agg	act	ggg	tct	tac	tac	tgg	act	tgg	gtt	cgc	cag	ccc	ccc	ggg	aag		250
Arg	Thr	Gly	Ser	Tyr	Tyr	Trp	Thr	Trp	Val	Arg	Gln	Pro	Pro	Gly	Lys		
30				35				40						45			
ggc	ctg	gag	tgg	att	ggc	tac	att	tat	tat	act	ggg	gac	acc	tac	tac		298
Gly	Leu	Glu	Trp	Ile	Gly	Tyr	Ile	Tyr	Tyr	Thr	Gly	Asp	Thr	Tyr	Tyr		
				50				55						60			
aac	ccg	tcc	ctc	aag	agt	cga	att	acc	atg	tcg	cta	gac	acg	tny	wag		346
Asn	Pro	Ser	Leu	Lys	Ser	Arg	Ile	Thr	Met	Ser	Leu	Asp	Thr	Xaa	Xaa		

38

```

      65      70      75
aac cag ttc kcc ctg agc ctg acc tct gtg act gtc gca gac acg g      392
Asn Gln Phe Xaa Leu Ser Leu Thr Ser Val Thr Val Ala Asp Thr
      80      85      90

```

<210> 53  
 <211> 172  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 14..172

<221> sig\_peptide  
 <222> 14..58  
 <223> Von Heijne matrix  
 score 10.8999996185303  
 seq LSVCLLLVTLALC/CY

```

<400> 53
aaaacaagcc acc atg aag ctg tcg gtg tgt ctc ctg ctg gtc acg ctg      49
              Met Lys Leu Ser Val Cys Leu Leu Leu Val Thr Leu
              -15              -10              -5
gcc ctc tgc tgc tac cag gcc aat gcc gag ttc tgc cca gct ctt gtt      97
Ala Leu Cys Cys Tyr Gln Ala Asn Ala Glu Phe Cys Pro Ala Leu Val
              1              5              10
tct gag ctg tta gac ttc ttc ttc att agt gaa cct ctg ttc aag tta      145
Ser Glu Leu Leu Asp Phe Phe Phe Ile Ser Glu Pro Leu Phe Lys Leu
              15              20              25
agt ctt gcc aaa ttt gat gcc cct cga      172
Ser Leu Ala Lys Phe Asp Ala Pro Arg
              30              35

```

<210> 54  
 <211> 259  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 190..258

<221> sig\_peptide  
 <222> 190..237  
 <223> Von Heijne matrix  
 score 10.8999996185303  
 seq VLLVLSLSQCCLS/DP

```

<400> 54
tacctggaaa gaacagaaat ttgttaatTT acaggtctga aggtgagaaa tctgaaatta      60
gtcttacaaa actaaaatga agttgttTga agccttgTct ccttctggag gttccagggg      120
aaaaaagtat gtttcctTga ctttccagcc kstacaggcc cacagcattc ctgcttgcag      180
ccctatgtc atg tca cct gtc ctc ttg gtg ctg tca ttg tca caa tgc ctt      231
              Met Ser Pro Val Leu Leu Val Leu Ser Leu Ser Gln Cys Leu
              -15              -10              -5
ctt tct gac cct gtc att cct ggc ctc c      259
Leu Ser Asp Pro Val Ile Pro Gly Leu
              1              5

```

<210> 55  
 <211> 320  
 <212> DNA



<213> Homo sapiens

<220>

<221> CDS

<222> 40..318

<221> sig\_peptide

<222> 40..96

<223> Von Heijne matrix

score 10.8999996185303

seq FLLLVAAPRWVLS/QV

<400> 55

```

aaacacsyyt tgagagtcct ggacctcctg tgcaggacc atg aaa cat ctg tgg      54
                                     Met Lys His Leu Trp
                                     -15
ttc ttc ctt ctc ctg gtg gca gct ccc aga tgg gtc ctg tcc cag gtg      102
Phe Phe Leu Leu Val Ala Ala Pro Arg Trp Val Leu Ser Gln Val
          -10                      -5                      1
cgg ctg cag gag tcg ggc cca cgg ctg gtg aag cct tcg gag amc ctg      150
Arg Leu Gln Glu Ser Gly Pro Arg Leu Val Lys Pro Ser Glu Xaa Leu
          5                      10                      15
tcc ctc acc tgt agt gtc tct ggt gtc tcc gtc act aat ttc ttc tgg      198
Ser Leu Thr Cys Ser Val Ser Gly Val Ser Val Thr Asn Phe Phe Trp
          20                      25                      30
aac tgg atc cgg aag ccc cca ggc aag ggc ctg gag tgg ctt ggc tac      246
Asn Trp Ile Arg Lys Pro Pro Gly Lys Gly Leu Glu Trp Leu Gly Tyr
          35                      40                      45                      50
atg tct tat ggc gtg agc aca aac tat cac ccc gcc tac cag agt cgg      294
Met Ser Tyr Gly Val Ser Thr Asn Tyr His Pro Ala Tyr Gln Ser Arg
          55                      60                      65
gtc agt ata tcg att gac acg tgg gg      320
Val Ser Ile Ser Ile Asp Thr Trp
          70

```

<210> 56

<211> 457

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 39..455

<221> sig\_peptide

<222> 39..95

<223> Von Heijne matrix

score 10.8999996185303

seq FLLLVAAPRWVLS/QV

<400> 56

```

aatactttct gagagtcctg gacctcctgt gcaagaac atg aaa cat ctg tgg ttc      56
                                     Met Lys His Leu Trp Phe
                                     -15
ttc ctt ctc ctg gtg gca gct ccc aga tgg gtc ctg tcc cag gtg cag      104
Phe Leu Leu Leu Val Ala Ala Pro Arg Trp Val Leu Ser Gln Val Gln
          -10                      -5                      1
ctg cag gag gcg ggc cca cga ctg gtg aag cct tcg gag gcc ctg tcc      152
Leu Gln Glu Ala Gly Pro Arg Leu Val Lys Pro Ser Glu Ala Leu Ser
          5                      10                      15
ctc acc tgc act gtc tct ggt gtc tcc agc agc aat tac gac tgg agt      200
Leu Thr Cys Thr Val Ser Gly Val Ser Ser Ser Asn Tyr Asp Trp Ser
          20                      25                      30                      35

```

40

tgg att cgg cag gcc cca ggg aag gga ctg gaa tgg att ggg tat ata	248
Trp Ile Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Ile Gly Tyr Ile	
40 45 50	
gac gat agt aag aat aga ggg agt acg acc tac aac ccc tcc ctc aag	296
Asp Asp Ser Lys Asn Arg Gly Ser Thr Thr Tyr Asn Pro Ser Leu Lys	
55 60 65	
agt cga gtc acc ata tcg stg gac acg tcc aag ast cag ttg tcc ctg	344
Ser Arg Val Thr Ile Ser Xaa Asp Thr Ser Lys Xaa Gln Leu Ser Leu	
70 75 80	
agg ctg acc tct gtg acc kcs gca gac acg gcc gtc tat tat tgt gcg	392
Arg Leu Thr Ser Val Thr Xaa Ala Asp Thr Ala Val Tyr Tyr Cys Ala	
85 90 95	
aga aag tca tct atg cat agt agt ggc tgg cat aac cgg agt ctc tac	440
Arg Lys Ser Ser Met His Ser Ser Gly Trp His Asn Arg Ser Leu Tyr	
100 105 110 115	
tgg tac ttc gat cct gg	457
Trp Tyr Phe Asp Pro	
120	

&lt;210&gt; 57

&lt;211&gt; 420

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 17..418

&lt;221&gt; sig\_peptide

&lt;222&gt; 17..94

&lt;223&gt; Von Heijne matrix

score 10.8999996185303

seq FLLLVAAPRWLS/QV

&lt;400&gt; 57

atactttctg agactc atg gac ctc ctg cac aag aac atg aaa gac ctg tgg	52
Met Asp Leu Leu His Lys Asn Met Lys Asp Leu Trp	
-25 -20 -15	
ttc ttc ctc ctc ctg gtg gca gct ccc aga tgg gtc ctg tct cag gtg	100
Phe Phe Leu Leu Leu Val Ala Ala Pro Arg Trp Val Leu Ser Gln Val	
-10 -5 1	
ctg cag gag tcg ggc cca gga ctg gtg aag cct tcg ggg acc ctg tcc	148
Leu Gln Glu Ser Gly Pro Gly Leu Val Lys Pro Ser Gly Thr Leu Ser	
5 10 15	
ctc acc tgc gct gtc tct ggt ggc tcc atc ata agt agt aat tgg tgg	196
Leu Thr Cys Ala Val Ser Gly Gly Ser Ile Ile Ser Ser Asn Trp Trp	
20 25 30	
agt tgg gtc cgc cag acc cca ggg aag ggg ctg gag tgg att ggg gaa	244
Ser Trp Val Arg Gln Thr Pro Gly Lys Gly Leu Glu Trp Ile Gly Glu	
35 40 45 50	
atc tat gaa gat ggg atc acc aac tac aac ccg tcc ctc aag agt cga	292
Ile Tyr Glu Asp Gly Ile Thr Asn Tyr Asn Pro Ser Leu Lys Ser Arg	
55 60 65	
gtc atc att tca gtg gac aag gcc aag aac cag ttc tcc ctg aag atg	340
Val Ile Ile Ser Val Asp Lys Ala Lys Asn Gln Phe Ser Leu Lys Met	
70 75 80	
agg tct gtg acc gcc tcg gac acg gcc gtc tat tac tgt gcg aga ggt	388
Arg Ser Val Thr Ala Ser Asp Thr Ala Val Tyr Tyr Cys Ala Arg Gly	
85 90 95	
agc agc tcg gtt cgg aca gac tac tgg ggc ca	420
Ser Ser Ser Val Arg Thr Asp Tyr Trp Gly	
100 105	

```
<210> 58
<211> 469
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 38..469
```

```
<221> sig_peptide
<222> 38..94
<223> Von Heijne matrix
      score 10.8999996185303
      seq  FLLLVAAPRWVLS/QV
```

<400>	58																	
atacttttctg agagtcctgg acctcctgtg caagaac atg aaa cac ctg tgg ttc																		55
Met Lys His Leu Trp Phe																		
-15																		
ttc	ctc	ctg	ctg	gtg	gca	gct	ccc	aga	tgg	gtc	ctg	tcc	cag	gtg	cag			103
Phe	Leu	Leu	Leu	Val	Ala	Ala	Pro	Arg	Trp	Val	Leu	Ser	Gln	Val	Gln			
-10																		
-5																		
1																		
ctg	cag	gag	tcc	ggg	tca	gga	ccg	gtg	gat	sct	tca	cag	acc	ctg	tsc			151
Leu	Gln	Glu	Ser	Gly	Ser	Gly	Pro	Val	Asp	Xaa	Xaa	Gln	Thr	Leu	Xaa			
5																		
10																		
15																		
ctc	acc	tgc	act	gks	tct	ggg	gtc	tcc	atc	agc	agt	agt	gat	aat	tgt			199
Leu	Thr	Cys	Thr	Xaa	Ser	Gly	Val	Ser	Ile	Ser	Ser	Ser	Asp	Asn	Cys			
20																		
25																		
30																		
35																		
tgg	agc	tgg	atc	cgg	cag	cca	cca	ggg	aag	ggc	ctg	gag	tgg	att	gga			247
Trp	Ser	Trp	Ile	Arg	Gln	Pro	Pro	Gly	Lys	Gly	Leu	Glu	Trp	Ile	Gly			
40																		
45																		
50																		
tac	atc	tat	cay	agt	ggg	ggg	acc	tac	tac	aac	ccg	acc	ctc	aag	agc			295
Tyr	Ile	Tyr	His	Ser	Gly	Gly	Thr	Tyr	Tyr	Asn	Pro	Thr	Leu	Lys	Ser			
55																		
60																		
65																		
cga	gtc	acc	atc	tgc	gba	gac	agg	atc	agg	aac	caa	ttc	tcc	ctg	aag			343
Arg	Val	Thr	Ile	Ser	Xaa	Asp	Arg	Ile	Arg	Asn	Gln	Phe	Ser	Leu	Lys			
70																		
75																		
80																		
ctg	agc	tct	gtg	acg	gcc	gag	gac	acg	gcc	gtg	tat	kac	tgt	ggc	aga			391
Leu	Ser	Ser	Val	Thr	Ala	Xaa	Asp	Thr	Ala	Val	Tyr	Xaa	Cys	Gly	Arg			
85																		
90																		
95																		
gca	cag	ggg	aga	atg	ggg	atc	ggg	acg	acg	att	ttt	gat	ctc	tgg	ggc			439
Ala	Gln	Gly	Arg	Met	Gly	Ile	Gly	Thr	Thr	Ile	Phe	Asp	Leu	Trp	Gly			
100																		
105																		
110																		
115																		
ggg	gga	caa	tgg	tca	ccg	tct	ctg	cag	cct									469
Gly	Gly	Gln	Trp	Ser	Pro	Ser	Leu	Gln	Pro									
120																		
125																		

```
<210> 59
<211> 471
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 52..471
```

```
<221> sig_peptide
<222> 52..108
<223> Von Heijne matrix
      score 10.8000001907349
      seq ILFLVAAATGAHS/QV
```

```
<221> misc_feature
```

&lt;222&gt; 210

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 59

```

acccaacaac cacatccctc ctcagmagcc cccagagcac aackcctyac c atg gac      57
                                     Met Asp
tgg acc tgg agg atc ctc ttt ttg gtg gca gca gcc aca ggt gcc cac      105
Trp Thr Trp Arg Ile Leu Phe Leu Val Ala Ala Ala Thr Gly Ala His
      -15                -10                -5
tcc cag gtc cag ctt gtg cag tct ggg gct gag gtg aag aag cct ggg      153
Ser Gln Val Gln Leu Val Gln Ser Gly Ala Glu Val Lys Lys Pro Gly
      1                5                10                15
gcc tca gtg aag gtt tcc tgc aag gct tct gga tac ayc ttc act ary      201
Ala Ser Val Lys Val Ser Cys Lys Ala Ser Gly Tyr Xaa Phe Thr Xaa
      20                25                30
tmt gct atn cat tgg gtg cgc cag gcc ccc gga car agr ctt gag tgg      249
Xaa Ala Xaa His Trp Val Arg Gln Ala Pro Gly Gln Arg Leu Glu Trp
      35                40                45
atg ggr tgg atc aac gct gcc amt ggt wam aca awa tat tca cag aas      297
Met Gly Trp Ile Asn Ala Ala Xaa Gly Xaa Thr Xaa Tyr Ser Gln Xaa
      50                55                60
ttc cag grc aga gtc acc wtt acc agg gac aca tcc gcg agc aca gtc      345
Phe Gln Xaa Arg Val Thr Xaa Thr Arg Asp Thr Ser Ala Ser Thr Val
      65                70                75
tcc atg gag ctg agc agc ctg aga tct gaa gac acg gct gtg tat ttc      393
Ser Met Glu Leu Ser Ser Leu Arg Ser Glu Asp Thr Ala Val Tyr Phe
      80                85                90                95
tgt gcg aga gat tgg gaa att gca gta gta cca act gct ata aac tct      441
Cys Ala Arg Asp Trp Glu Ile Ala Val Val Pro Thr Ala Ile Asn Ser
      100                105                110
tac ggg ttc gac cct ggg gcc agg gaa cct      471
Tyr Gly Phe Asp Pro Gly Ala Arg Glu Pro
      115                120

```

&lt;210&gt; 60

&lt;211&gt; 348

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 193..348

&lt;221&gt; sig\_peptide

&lt;222&gt; 193..270

&lt;223&gt; Von Heijne matrix

score 10.8000001907349

seq VLFLCVFLGMSWA/GA

&lt;400&gt; 60

```

agagcaaaga ggcaatctga agagaaaagc ataggaaagg aaacagtggg aataggaatt      60
ggggtaaaat gaggatcctt cccacaaaac attgctatta ttcagctcat ttcaaaggat      120
tccgstgcwg ccatttggtga gagccgctgg aggctgagtg aaagtcattt tgaaagactg      180
atccaaagaa ga atg gag gcc aga gtg gag cgt gct gtg cag aaa agg caa      231
      Met Glu Ala Arg Val Glu Arg Ala Val Gln Lys Arg Gln
      -25                -20                -15
gtc tta ttt ctt tgt gta ttt ctg gga atg tct tgg gct ggc gcc gaa      279
Val Leu Phe Leu Cys Val Phe Leu Gly Met Ser Trp Ala Gly Ala Glu
      -10                -5                1
ccg ctt cgg tat ttt gtg gcg gag gaa acc gag aga ggc acc tdk ctt      327
Pro Leu Arg Tyr Phe Val Ala Glu Glu Thr Glu Arg Gly Thr Xaa Leu
      5                10                15

```

acc aac ttg gca aaa gac cta  
 Thr Asn Leu Ala Lys Asp Leu  
 20 25

348

<210> 61  
 <211> 457  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 55..456

<221> sig\_peptide  
 <222> 55..111  
 <223> Von Heijne matrix  
 score 10.8000001907349  
 seq ILFLVAAATGAHS/QV

<400> 61  
 acccaaaaac cacacccctc cttgggagaa tcccctagat cacagctcct cacc atg 57  
 Met  
 gac tgg acc tgg agc atc ctt ttc ttg gtg gca gca gcg aca ggt gcc 105  
 Asp Trp Thr Trp Ser Ile Leu Phe Leu Val Ala Ala Ala Thr Gly Ala  
 -15 -10 -5  
 cac tcc cag gtt cag ctg gtg cag tct gga ggt gag gtg aag aag cct 153  
 His Ser Gln Val Gln Leu Val Gln Ser Gly Gly Glu Val Lys Lys Pro  
 1 5 10  
 ggg gcc tcc gtc aag gtc tcc tgc aag gct tct ggt tac acc ttt acc 201  
 Gly Ala Ser Val Lys Val Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr  
 15 20 25 30  
 aga tat gat atc aac tgg gtg cga cag gcc cct gga caa ggg ctt gag 249  
 Arg Tyr Asp Ile Asn Trp Val Arg Gln Ala Pro Gly Gln Gly Leu Glu  
 35 40 45  
 tgg atg gga tgg atc agc gct dcc aat ggt aac aca aat tat gca cag 297  
 Trp Met Gly Trp Ile Ser Ala Xaa Asn Gly Asn Thr Asn Tyr Ala Gln  
 50 55 60  
 daa gtc cag ggc aga gtc acc atg acc aca gac aca tcc acg aga aca 345  
 Xaa Val Gln Gly Arg Val Thr Met Thr Thr Asp Thr Ser Thr Arg Thr  
 65 70 75  
 gcc tac atg gaa ctg agg agc ctg cga tct gac gac acg gcc att tat 393  
 Ala Tyr Met Glu Leu Arg Ser Leu Arg Ser Asp Asp Thr Ala Ile Tyr  
 80 85 90  
 tac tgt gcg cga gag atm bta gtg gba sta tgt gat gga cag ttg ggg 441  
 Tyr Cys Ala Arg Glu Ile Xaa Val Xaa Xaa Cys Asp Gly Gln Leu Gly  
 95 100 105 110  
 cca ggg aac ctg gtc a  
 Pro Gly Asn Leu Val 457  
 115

<210> 62  
 <211> 439  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 18..437

<221> sig\_peptide  
 <222> 18..95  
 <223> Von Heijne matrix  
 score 10.8000001907349

seq FLLVAAPRWVLS/QE

[illegible]

```
<210> 63
<211> 214
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 82..213  
  
<221> sig_peptide  
<222> 82..126  
<223> Von Heijne matrix  
score 10.6999998092651  
seq LLALFFLLRIALA/SQ
```

<400> 63		
accattgggtg tgtctgtttt tatgccagta ctgtgatgtt ttggttatat agctttgtaa		60
tatatatttga agccagatag t atg atg ctt cta gct ttg ttc ttt ttg ctt		111
	Met Met Leu Leu Ala Leu Phe Phe Leu Leu	
	-15 -10	
agg att gct ttg gct agt caa ggt ctt ttg tgg ttc cat aca aat ttt		159
Arg Ile Ala Leu Ala Ser Gln Gly Leu Leu Trp Phe His Thr Asn Phe		
-5 1 5 10		
aag gtt ttt gtt gtt tcy att tgt gtg aag act atc att ggg att tcg		207
Lys Val Phe Val Val Ser Ile Cys Val Lys Thr Ile Ile Gly Ile Ser		
	15 20 25	
ggg ggc a		214
Gly Gly		
<210> 64		

<211> 297  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 63..296

<221> sig\_peptide  
 <222> 63..119  
 <223> Von Heijne matrix  
 score 10.6999998092651  
 seq ILFLVAAATGALS/QV

<400> 64  
 gtgcatcacc cagcaaccac atctgtcctc tagagaatcc cctgagadht cggttcctca 60  
 cc atg gac tgg acc tgg agg atc ctc ttc ttg gtg gca gcr gcc aca 107  
 Met Asp Trp Thr Trp Arg Ile Leu Phe Leu Val Ala Ala Ala Thr  
 -15 -10 -5  
 gga gcc ctc tcc cag gtg cag ctg gtr cag tct gga ggt gar gtg aag 155  
 Gly Ala Leu Ser Gln Val Gln Leu Val Gln Ser Gly Gly Glu Val Lys  
 1 5 10  
 aag cct ggg gcc tca gtg agg gtc tcc tgc aag gcc tct gga tac agc 203  
 Lys Pro Gly Ala Ser Val Arg Val Ser Cys Lys Ala Ser Gly Tyr Ser  
 15 20 25  
 ttc atc ggc tat tat gta cac tgg ata cga cag act cct ggg cga sgc 251  
 Phe Ile Gly Tyr Tyr Val His Trp Ile Arg Gln Thr Pro Gly Arg Xaa  
 30 35 40  
 ctt gag tgg atg ggg tgg gtc aac cct crs act ggc gac aac ggg g 297  
 Leu Glu Trp Met Gly Trp Val Asn Pro Xaa Thr Gly Asp Asn Gly  
 45 50 55

<210> 65  
 <211> 370  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 237..368

<221> sig\_peptide  
 <222> 237..347  
 <223> Von Heijne matrix  
 score 10.6000003814697  
 seq YLLLVLSLSLCS/CS

<400> 65  
 aaaggtacac aattgaaaaa aattgtatcc ttcacaacag atgtgggag tcaactttta 60  
 gacctgtgtg ctttagtttg acctgtcctt cagtgagtgt atataaaatt ctaagctaaa 120  
 acatatatttc tgaattgtg aaggattatgc atgtctatct tcttgacctac tctaaatata 180  
 tcaatcgttt tcttggaag ttagtccttc tttcacactt gtctgtagat ctttac atg 239  
 Met  
 ttc ttt cag ttt tgg aag tcc tct gca tat tta ata ttt gtt agt att 287  
 Phe Phe Gln Phe Trp Lys Ser Ser Ala Tyr Leu Ile Phe Val Ser Ile  
 -35 -30 -25  
 tgt aaa ggt ttt ctt cct gtc tac ctc ctt ctt gtt ctc tct ctc tct 335  
 Cys Lys Gly Phe Leu Pro Val Tyr Leu Leu Leu Val Leu Ser Leu Ser  
 -20 -15 -10 -5  
 ctc tct ctc tgt tgc tct ctc ttg ctc tct ctc ca 370  
 Leu Ser Leu Cys Cys Ser Leu Leu Leu Ser Leu  
 1 5

<210> 66  
 <211> 428  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 45..428

<221> sig\_peptide  
 <222> 45..101  
 <223> Von Heijne matrix  
       score 10.6000003814697  
       seq ILFLVAAATGVHS/QV

<221> misc\_feature  
 <222> 342..343  
 <223> n=a, g, c or t

```

<400> 66
aaccacmycc ctctcagaa gccccagag cacaacgcct cacc atg gac tgg acc      56
                                     Met Asp Trp Thr
tgg agg atc ctc ttt ttg gtg gca gca gcc aca ggt gtc cac tcc cag      104
Trp Arg Ile Leu Phe Leu Val Ala Ala Ala Thr Gly Val His Ser Gln
-15                               -10          -5              1
gtc cac ctt gtt cag tct ggg gct gar gtg aag aag cct ggg act ccg      152
Val His Leu Val Gln Ser Gly Ala Glu Val Lys Lys Pro Gly Thr Pro
                    5              10              15
gtg aac att tcc tgt aag gct ttt ggc tac acc ttc cct gcc ttt gct      200
Val Asn Ile Ser Cys Lys Ala Phe Gly Tyr Thr Phe Pro Ala Phe Ala
                20              25              30
ata cat tgg gtt cgc cag gcc ccc gga caa agt ctt gag tgg atg gga      248
Ile His Trp Val Arg Gln Ala Pro Gly Gln Ser Leu Glu Trp Met Gly
                35              40              45
tgg gtc aac att ggc cat ggc aac aca aag tat tca cag aag ttt cag      296
Trp Val Asn Ile Gly His Gly Asn Thr Lys Tyr Ser Gln Lys Phe Gln
50              55              60              65
ggc aga ctc gcc atc tcc aga gac acg tcc gcg aac ata gtc tac nng      344
Gly Arg Leu Ala Ile Ser Arg Asp Thr Ser Ala Asn Ile Val Tyr Xaa
                70              75              80
gaa ctg agc ggc ctg aga tct gaa gac acg gct gtc tat tac tgt gcg      392
Glu Leu Ser Gly Leu Arg Ser Glu Asp Thr Ala Val Tyr Tyr Cys Ala
                85              90              95
agg gat aat ctt ttc ttt ggc agt atg ggc ttt gac      428
Arg Asp Asn Leu Phe Phe Gly Ser Met Gly Phe Asp
                100              105

```

<210> 67  
 <211> 493  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 38..493

<221> sig\_peptide  
 <222> 38..85  
 <223> Von Heijne matrix  
       score 10.6000003814697  
       seq TVLLGLLSHCTG/SV



```
<210> 68
<211> 180
<212> DNA
<213> Homo sapiens
```

$\langle 220 \rangle$

<221> CDS

<222> 36...179

<221> sig peptide

<222> 36.80

```
<223> Von Heijne matrix
      score 10.6000003814697
      seq LLFLLLFVCFSRQ/GL
```

<400> 68

tggcagttac tccagctccc aaatatagat attcc	atg agg ttg ttg ttt ttg	53
	Met Arg Leu Leu Phe Leu	
	-15 -10	
ttg ttg ttt gtt tgt ttt tcg aga cag ggt ctc gct ttg tct ctc agg		101
Leu Leu Phe Val Cys Phe Ser Arg Gln Gly Leu Ala Leu Ser Leu Arg		
	-5 1 5	
ctg gaa tgc agt ggt atg atc atg gct tac tgc agc atc agc ctc cca		149
Leu Glu Cys Ser Gly Met Ile Met Ala Tyr Cys Ser Ile Ser Leu Pro		
	10 15 20	
ggc tca agc agt cct ctc acc tca gcc tcc a		180
Gly Ser Ser Ser Pro Leu Thr Ser Ala Ser		
	25 30	

<210> 69  
 <211> 259  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 38..259

<221> sig\_peptide  
 <222> 38..94  
 <223> Von Heijne matrix  
 score 10.6000003814697  
 seq FLLLVSA PRWVLS/QV

<400> 69  
 atacttyctg agagtcctgg acctcctgca caagaac atg aaa cac ctg tgg ttc 55  
 Met Lys His Leu Trp Phe  
 -15  
 ttc ctc ctc ctg gtg tca gct ccc aga tgg gtc ctg tct cag gtg cag 103  
 Phe Leu Leu Leu Val Ser Ala Pro Arg Trp Val Leu Ser Gln Val Gln  
 -10 -5 1  
 cta cag gag tcg ggc cca gga ctg gtg aag cct tcg ggg agg ctg tcc 151  
 Leu Gln Glu Ser Gly Pro Gly Leu Val Lys Pro Ser Gly Arg Leu Ser  
 5 10 15  
 ctc gcc tgc gat gtg gtg gaa ttg agt ccg ccg gcc ccc agg ggc ggg 199  
 Leu Ala Cys Asp Val Val Glu Leu Ser Pro Pro Ala Pro Arg Gly Gly  
 20 25 30 35  
 tct gca gtg cat ctc aga aat ctt tca tca tgg gag ccc cac cta caa 247  
 Ser Ala Val His Leu Arg Asn Leu Ser Ser Trp Glu Pro His Leu Gln  
 40 45 50  
 ccc gtc tcg ggg 259  
 Pro Val Ser Gly  
 55

<210> 70  
 <211> 178  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 7..177

<221> sig\_peptide  
 <222> 7..102  
 <223> Von Heijne matrix  
 score 10.6000003814697  
 seq VVFLLLLVSTLSS/VV

<400> 70  
 cgtata atg act tac ttt cct ctg ggt aga tac cca gta atg gga ttg 48  
 Met Thr Tyr Phe Pro Leu Gly Arg Tyr Pro Val Met Gly Leu  
 -30 -25 -20  
 ctg gat caa atg gta gtt gtg ttt tta ctt ctt tta gtc tcc aca ctt 96  
 Leu Asp Gln Met Val Val Val Phe Leu Leu Leu Leu Val Ser Thr Leu  
 -15 -10 -5  
 tct tcc gta gtg gtt tta cta gtt tgc att ccc acc agc agt gta aaa 144  
 Ser Ser Val Val Val Leu Leu Val Cys Ile Pro Thr Ser Ser Val Lys  
 1 5 10  
 ttg ttc cct ttt cac cat atc cac acc aac tgg g 178  
 Leu Phe Pro Phe His His Ile His Thr Asn Trp

49

```

15                      20                      25

<210> 71
<211> 131
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 40..129

<221> sig_peptide
<222> 40..96
<223> Von Heijne matrix
      score 10.5
      seq WVLLVAMLRGLQC/QV

<400> 71
agctctggga gacgagccca gctctgcagt ggactcacc atg gag ttt ggg ctg      54
                                   Met Glu Phe Gly Leu
                                   -15
agc tgg gtt ctc ctc gtt gct atg tta aga ggt ctc cag tgt caa gtg      102
Ser Trp Val Leu Leu Val Ala Met Leu Arg Gly Leu Gln Cys Gln Val
      -10                      -5                      1
cag ctg gtg gag tct ggg gga acc gcg gg      131
Gln Leu Val Glu Ser Gly Gly Thr Ala
      5                      10

<210> 72
<211> 217
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 47..217

<221> sig_peptide
<222> 47..91
<223> Von Heijne matrix
      score 10.5
      seq LSLILLENVSG/FP

<400> 72
ttgcttacaa ttttaatgtg tctcattgct actggtcctc cttcta atg tat ctg      55
                                   Met Tyr Leu
                                   -15
agc ttg tta att cta ctt ttg gaa aat gtc agt ggc ttt ccc ttt cct      103
Ser Leu Leu Ile Leu Leu Leu Glu Asn Val Ser Gly Phe Pro Phe Pro
      -10                      -5                      1
cta att ttc cag ctt cat gca tcc cct ggc cat aag ata ctt cca gac      151
Leu Ile Phe Gln Leu His Ala Ser Pro Gly His Lys Ile Leu Pro Asp
      5                      10                      15                      20
tgt atg ata tat tct atc act gtc agc ctt atg ttc cct gtg gtt gac      199
Cys Met Ile Tyr Ser Ile Thr Val Ser Leu Met Phe Pro Val Val Asp
      25                      30                      35
tat ata agc acg caa ggg      217
Tyr Ile Ser Thr Gln Gly
      40

<210> 73
<211> 192
<212> DNA

```

<213> Homo sapiens

<220>

<221> CDS

<222> 100..192

<221> sig\_peptide

<222> 100..183

<223> Von Heijne matrix

score 10.5

seq SLWFXCLLFLFA/WP

<400> 73

agttaaaatc atgtactgtg atcagtcacc tgggtttttga tttttatgaa gggtttttttt 60

gttttagatag ttgttaaatt tgggtgttcct gtggggagg atg atg aga gcc ttc 114

Met Met Arg Ala Phe

-25

tat ttg gct atc ttg ttc tgc ctc tct ctc tcc tta tgg ttc tdk tgt 162

Tyr Leu Ala Ile Leu Phe Cys Leu Ser Leu Ser Leu Trp Phe Xaa Cys

-20

-15

-10

tta ctt ttt ttg ctt ttt gct tgg cct ggg 192

Leu Leu Phe Leu Leu Phe Ala Trp Pro Gly

-5

1

<210> 74

<211> 329

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 23..328

<221> sig\_peptide

<222> 23..82

<223> Von Heijne matrix

score 10.3999996185303

seq FLTLLHCTGSLA/QL

<400> 74

agagctctgg ggagctctgca cc atg gct tgg acc cca ctc ctc ttc ctc acc 52

Met Ala Trp Thr Pro Leu Leu Phe Leu Thr

-20

-15

ctc ctc ctc cac tgc aca ggg tct ctc gcc cag ctt gtg ctg act caa 100

Leu Leu Leu His Cys Thr Gly Ser Leu Ala Gln Leu Val Leu Thr Gln

-10

-5

1

5

tcg ccc tct gcc tct gcc tcc ctg gga gcc tcg gtc aag ctc acc tgc 148

Ser Pro Ser Ala Ser Ala Ser Leu Gly Ala Ser Val Lys Leu Thr Cys

10

15

20

act ctg agc agt ggg cac agc aac tac ggc atc gct tgg tat cag cag 196

Thr Leu Ser Ser Gly His Ser Asn Tyr Gly Ile Ala Trp Tyr Gln Gln

25

30

35

cag cca gag aag ggc cct cga ttc ttg atg aaa gtt aac agt gat ggc 244

Gln Pro Glu Lys Gly Pro Arg Phe Leu Met Lys Val Asn Ser Asp Gly

40

45

50

agc cac atg aag gcg gac ggg atc cct gat cgc ttc tca ggc tcc agc 292

Ser His Met Lys Ala Asp Gly Ile Pro Asp Arg Phe Ser Gly Ser Ser

55

60

65

70

tct ggg gct gag cgc tac ctc tcc atc tcc agc ctc a 329

Ser Gly Ala Glu Arg Tyr Leu Ser Ile Ser Ser Leu

75

80

<210> 75

<211> 314  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 259..312

<221> sig\_peptide  
 <222> 259..300  
 <223> Von Heijne matrix  
 score 10.3999996185303  
 seq PLALFFLLSVALA/IQ

<400> 75  
 tagggtgaga gatggggatc tagtttttatt cttctgcata tggatatcca gttttcccag 60  
 taacatttat tgaagagact ggcctttccc caatgagtggt tcttggcacc tttgtcaaaa 120  
 gtcagttggc cgtagatatg tggattaatt tctgtgttcc ctgttttggt ccattggcct 180  
 atgtgtctgt ttttatgaca gtaccagggt gttttgggta ctacagcttt gtagtttact 240  
 ttgaggtctg ttagtggtg atg cct cta gct ttg ttc ttt ttg ctc agt gtt 291  
 Met Pro Leu Ala Leu Phe Phe Leu Leu Ser Val  
 -10 -5  
 gct ttg gct att cag ggt cag gg 314  
 Ala Leu Ala Ile Gln Gly Gln  
 1

<210> 76  
 <211> 447  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 59..445

<221> sig\_peptide  
 <222> 59..115  
 <223> Von Heijne matrix  
 score 10.3999996185303  
 seq XFCLLAVAPGAHS/QV

<400> 76  
 atcatccaac aaccacatcc cttctctaca gaagcctctg agaggaaagt tcttcacc 58  
 atg gac tgg acc tgg agg rwc ttc tgc ttg ctg gct gta gct cca ggt 106  
 Met Asp Trp Thr Trp Arg Xaa Phe Cys Leu Leu Ala Val Ala Pro Gly  
 -15 -10 -5  
 gct cac tcc cag gtg cag ctg gtg cag tct ggg gct gag gtg aag aag 154  
 Ala His Ser Gln Val Gln Leu Val Gln Ser Gly Ala Glu Val Lys Lys  
 1 5 10  
 cct ggg gcc tca gtg aag gtt tcc tgc aag gca tct gga tac acc ttc 202  
 Pro Gly Ala Ser Val Lys Val Ser Cys Lys Ala Ser Gly Tyr Thr Phe  
 15 20 25  
 acc agc cac tat atg cac tgg gtg cga cag gcc cct gga caa ggg ctt 250  
 Thr Ser His Tyr Met His Trp Val Arg Gln Ala Pro Gly Gln Gly Leu  
 30 35 40 45  
 gag tgg atg gga ata atc tac cct gat agt gat acc act aag tac cba 298  
 Glu Trp Met Gly Ile Ile Tyr Pro Asp Ser Asp Thr Thr Lys Tyr Xaa  
 50 55 60  
 cag aac ttc cag ggc aga gtc acc atg act agg gac acg tcc acg agc 346  
 Gln Asn Phe Gln Gly Arg Val Thr Met Thr Arg Asp Thr Ser Thr Ser  
 65 70 75  
 aca gtc tac atg gag ctg agc agc ctg aca tct gac gac acg gcc gtg 394  
 Thr Val Tyr Met Glu Leu Ser Ser Leu Thr Ser Asp Asp Thr Ala Val

52

80	85	90	
tat tat tgt gct aga gag gcg tat agt ggg agc tac cgc ttt gac tac	442		
Tyr Tyr Cys Ala Arg Glu Ala Tyr Ser Gly Ser Tyr Arg Phe Asp Tyr			
95	100	105	
tgg gg			447
Trp			
110			

<210> 77  
 <211> 388  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 16..387

<221> sig\_peptide  
 <222> 16..93  
 <223> Von Heijne matrix  
 score 10.3000001907349  
 seq LLLLVAAPRWLS/QL

<400> 77	
agctttctga gagtc atg gat ctc atg tgc aag aaa atg aga cac ctg tgg	51
Met Asp Leu Met Cys Lys Lys Met Arg His Leu Trp	
-25 -20 -15	
ttc ctc ctc ctg ctg gtg gcg gct ccc aga tgg gtc ctg tcc cag ctg	99
Phe Leu Leu Leu Val Ala Ala Pro Arg Trp Val Leu Ser Gln Leu	
-10 -5 1	
cag ctt cag gag tcg ggc cca gga ctg gtg aag gct tcg gag acc ctg	147
Gln Leu Gln Glu Ser Gly Pro Gly Leu Val Lys Ala Ser Glu Thr Leu	
5 10 15	
tcc ctc gcc tgc agt gtc tct ggt gac tcc atc agc agt ggt aat tat	195
Ser Leu Ala Cys Ser Val Ser Gly Asp Ser Ile Ser Ser Gly Asn Tyr	
20 25 30	
tac tgg ggc tgg atc cgg cag ccc cca ggg aag gga ctg cag tgg ctt	243
Tyr Trp Gly Trp Ile Arg Gln Pro Pro Gly Lys Gly Leu Gln Trp Leu	
35 40 45 50	
ggg agt ctt tgg aat cgt ggc ggt ccg caa tac aay hcc tcc ctc aag	291
Gly Ser Leu Trp Asn Arg Gly Gly Pro Gln Tyr Asn Xaa Ser Leu Lys	
55 60 65	
aat cga gtc acc gtg tcc gta gac acg tcc acg aat cat ttc ttt ctg	339
Asn Arg Val Thr Val Ser Val Asp Thr Ser Thr Asn His Phe Phe Leu	
70 75 80	
aga ctg aat tcc gtg aay vgh gga cac ggc aat tta tta ctg tgc gcg a	388
Arg Leu Asn Ser Val Asn Xaa Gly His Gly Asn Leu Leu Leu Cys Ala	
85 90 95	

<210> 78  
 <211> 121  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 25..120

<221> sig\_peptide  
 <222> 25..72  
 <223> Von Heijne matrix  
 score 10.1999998092651  
 seq XLXLSVLLGXXXX/KX

<400> 78  
ctgaccttcc ctcccggcag cagc atg cgg grg tkg ctg kys ttg agt gtc 51  
Met Arg Xaa Xaa Leu Xaa Leu Ser Val  
-15 -10

ctg ttr ggg sck rts tbt kgc aag gmg gac ttt gtg ggg cat cag gtg 99  
Leu Leu Gly Xaa Xaa Xaa Xaa Lys Xaa Asp Phe Val Gly His Gln Val  
-5 1 5

ctc cga atc tct gta gcc gat g 121  
Leu Arg Ile Ser Val Ala Asp  
10 15

<210> 79  
<211> 149  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 36..149

<221> sig\_peptide  
<222> 36..143  
<223> Von Heijne matrix  
score 10.1999998092651  
seq FLLFFCFVFCLRG/QG

<400> 79  
tccgagcagc cagctctgtg taagcacatc cagga atg gca gaa tcc agg gag 53  
Met Ala Glu Ser Arg Glu  
-35

gaa ggt gaa agc tgt gtt gag agc cac tgt gtg ctc ttt ttc acc ctg 101  
Glu Gly Glu Ser Cys Val Glu Ser His Cys Val Leu Phe Phe Thr Leu  
-30 -25 -20 -15

ttt ttt ttg ttg ttt ttt tgt ttt gtt ttt tgt ttg agg gga cag ggg 149  
Phe Phe Leu Leu Phe Phe Cys Phe Val Phe Cys Leu Arg Gly Gln Gly  
-10 -5 1

<210> 80  
<211> 410  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 80..409

<221> sig\_peptide  
<222> 80..136  
<223> Von Heijne matrix  
score 10.1000003814697  
seq WLFLVAFLKGVQC/EV

<400> 80  
agctctgaga gaggagccca gccctgggat ttccaggtgt ttccatgtgg tgatcaggac 60  
tgaacagaga gaactcacc atg gag ctt ggg ctg agc tgg ctt ttt ctt gtg 112  
Met Glu Leu Gly Leu Ser Trp Leu Phe Leu Val  
-15 -10

gct ttt tta aaa ggt gtc cag tgt gag gtg cag ttg ttg gag tct ggg 160  
Ala Phe Leu Lys Gly Val Gln Cys Glu Val Gln Leu Leu Glu Ser Gly  
-5 1 5

gga ggc ttg gtc cag cct ggg ggg tcc ctg aga ctc tca tgt gca gcc 208  
Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala

54

10	15	20	
tcc gga ttc acc ttt agc tcc tat gcc atg ctc tgg gtc cgc cag gct			256
Ser Gly Phe Thr Phe Ser Ser Tyr Ala Met Leu Trp Val Arg Gln Ala			
25	30	35	40
cca ggt aag ggg ctg gag tgg gtc tca ggt att agt gct ggt gct gat			304
Pro Gly Lys Gly Leu Glu Trp Val Ser Gly Ile Ser Ala Gly Ala Asp			
45	50	55	
gat aca tat gat gca gac tcc gtg aag ggc cgg ttc acc att tcc aga			352
Asp Thr Tyr Asp Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg			
60	65	70	
gac gat tcc aag aaa atc cta tat cta caa atg aac agc ctg aga gcc			400
Asp Asp Ser Lys Lys Ile Leu Tyr Leu Gln Met Asn Ser Leu Arg Ala			
75	80	85	
gag gac agg c			410
Glu Asp Arg			
90			

<210> 81  
 <211> 219  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 38..217

<221> sig\_peptide  
 <222> 38..106  
 <223> Von Heijne matrix  
 score 10.1000003814697  
 seq VLGLLVFLTCYA/DD

<400> 81	
gaacaattta tctgcacgaa taccctgtgc taccaga atg gct gtc tca gta ctt	55
	Met Ala Val Ser Val Leu
	-20
cgc ctg aca gtt gtc ctg gga ctg ctt gtc tta ttc ctg acc tgc tat	103
Arg Leu Thr Val Val Leu Gly Leu Leu Val Leu Phe Leu Thr Cys Tyr	
-15	-10
gca gac gac aaa cca gac aag cca gac gac aag cca gac gac tgc ggc	151
Ala Asp Asp Lys Pro Asp Lys Pro Asp Asp Lys Pro Asp Asp Ser Gly	
1	5
aaa gac cca aag cca gac ttc ccc aaa ttc cta agc ctc ctg ggc aca	199
Lys Asp Pro Lys Pro Asp Phe Pro Lys Phe Leu Ser Leu Leu Gly Thr	
20	25
gag atc att gag aat gcg gg	219
Glu Ile Ile Glu Asn Ala	
35	

<210> 82  
 <211> 399  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 81..398

<221> sig\_peptide  
 <222> 81..152  
 <223> Von Heijne matrix  
 score 10  
 seq LLLLQALPSPLSA/RA



```

<400> 82
gaagaagagg gtagaggagg agagggagga ggaggaggga ggtggcgcg ccgtggcgga      60
ggagcaggag caggaggggg atg gag agg aga agg ctc ctg ggt ggc atg gcg      113
                Met Glu Arg Arg Arg Leu Leu Gly Gly Met Ala
                        -20                                -15
ctc ctg ctc ctc cag gcg ctg ccc agc ccc ttg tca gcc agg gct gaa      161
Leu Leu Leu Leu Gln Ala Leu Pro Ser Pro Leu Ser Ala Arg Ala Glu
                -10                                -5                                1
ccc ccg cag gat aag gaa gcc tgt gtg ggt acc aac aat caa agc tac      209
Pro Pro Gln Asp Lys Glu Ala Cys Val Gly Thr Asn Asn Gln Ser Tyr
                5                                10                                15
atc tgt gac aca gga cac tgc tgt gga cag tct cag tgc tgy aac tac      257
Ile Cys Asp Thr Gly His Cys Cys Gly Gln Ser Gln Cys Cys Asn Tyr
20                25                30                35
tac tat gaa ctc tgg tgg ttc tgg ctg gtg tgg acc atc atc atc atc      305
Tyr Tyr Glu Leu Trp Trp Phe Trp Leu Val Trp Thr Ile Ile Ile Ile
                40                45                50
ctg agc tgc tgc tgt gtt tgc cac cac cgc cga gcc aag cac cgc ctt      353
Leu Ser Cys Cys Cys Val Cys His His Arg Arg Ala Lys His Arg Leu
                55                60                65
cag gcc cag cag cgg caa cat gaa atc aac ctg atc gct tac cga g      399
Gln Ala Gln Gln Arg Gln His Glu Ile Asn Leu Ile Ala Tyr Arg
                70                75                80

```

```

<210> 83
<211> 398
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> CDS
<222> 288..398

<221> sig_peptide
<222> 288..368
<223> Von Heijne matrix
      score 9.89999961853027
      seq LCLLLFSLSLFLC/HE

```

```

<400> 83
cactctacct ctgacagcat gtatatgca ccagtagcta acaaaaactg gtctagtcaa      60
accaaattggg cacaaaagaa ccaggatacc aaaagttaag ctcatacagc tgcaaaccat      120
atcacttctt ggtaacaatg caagacctca taaacctaaa gaagagaaag aaaagaaaac      180
ttttgttact ttvctttttt gcttgtcact tatatacagg ctatgtgaga atataatttg      240
taggtataac acattaagaa aaagttatct tcattggata gaattga atg gtg gtc      296
                Met Val Val
                        -25
gct gat agg aat agg gcg tcc tct agc tct tat ctc tgt ctc tta ctc      344
Ala Asp Arg Asn Arg Ala Ser Ser Ser Ser Tyr Leu Cys Leu Leu Leu
                -20                -15                -10
ttt tct ctt tct ctt ttt ctc tgt cat gag act gtg tgt gac agg gcc      392
Phe Ser Leu Ser Leu Phe Leu Cys His Glu Thr Val Cys Asp Arg Ala
                -5                1                5
acc tgt
Thr Cys
      10

```

```

<210> 84
<211> 488
<212> DNA
<213> Homo sapiens

```

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 62..487

&lt;221&gt; sig\_peptide

&lt;222&gt; 62..118

&lt;223&gt; Von Heijne matrix

score 9.89999961853027

seq FLFVVAAATGVQS/QV

&lt;221&gt; misc\_feature

&lt;222&gt; 210,293

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 84

```

agcatcacat aacaaccaca ttcctcctct aaagaagccc ctgggagcac agctcatcac      60
c atg gac tgg acc tgg agg ttc ctc ttt gtg gtg gca gca gct aca ggt      109
  Met Asp Trp Thr Trp Arg Phe Leu Phe Val Val Ala Ala Ala Thr Gly
      -15          -10          -5
gtc cag tcc cag gtg cag ctg gtg cag tct ggg gct gag gtg aag aag      157
Val Gln Ser Gln Val Gln Leu Val Gln Ser Gly Ala Glu Val Lys Lys
      1          5          10
cct ggg tcc tcg gtg aag gtc tcc tgc aag gct tct gga ggc acc ttc      205
Pro Gly Ser Ser Val Lys Val Ser Cys Lys Ala Ser Gly Gly Thr Phe
      15          20          25
agc anc tat gct atc agc tgg gtg cga cag gcc cct gga caa ggg ctt      253
Ser Xaa Tyr Ala Ile Ser Trp Val Arg Gln Ala Pro Gly Gln Gly Leu
      30          35          40          45
gag tgg atg gga ggg atc atc cct atc ttt ggt aca gca nac tac gca      301
Glu Trp Met Gly Gly Ile Ile Pro Ile Phe Gly Thr Ala Xaa Tyr Ala
      50          55          60
cag aag ttc cag ggc aga gtc acs att acc gcg gac gra tcc acg asc      349
Gln Lys Phe Gln Gly Arg Val Thr Ile Thr Ala Asp Xaa Ser Thr Xaa
      65          70          75
aca rcc tac atg gag ctg agc agc ctg aga tct gag gac acg gcc stg      397
Thr Xaa Tyr Met Glu Leu Ser Ser Leu Arg Ser Glu Asp Thr Ala Xaa
      80          85          90
tat tac tgt gcg aga ggt caa gcc ccc ggt agg gta gta gta cca ctt      445
Tyr Tyr Cys Ala Arg Gly Gln Ala Pro Gly Arg Val Val Val Pro Leu
      95          100          105
ttc ctc tgg ggc cag gga acc tgg tca ccg tct cct cag cct c      488
Phe Leu Trp Gly Gln Gly Thr Trp Ser Pro Ser Pro Gln Pro
      110          115          120

```

&lt;210&gt; 85

&lt;211&gt; 290

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 30..290

&lt;221&gt; sig\_peptide

&lt;222&gt; 30..164

&lt;223&gt; Von Heijne matrix

score 9.89999961853027

seq LLSLLSFLDETSG/LS

&lt;400&gt; 85

```

cttcttttttc ttcgtaactt catggcaac atg acc tac agt tac tca ttt ttc      53
  Met Thr Tyr Ser Tyr Ser Phe Phe

```

```

                                -45                -40
agg cct gag ttg atc gtt aat cat ctt aat tat gtt cat tct gaa gcc      101
Arg Pro Glu Leu Ile Val Asn His Leu Asn Tyr Val His Ser Glu Ala
      -35                -30                -25
aac agg aga acc aag acc aaa act tta ttg tct ctg ctt tca ttt ctt      149
Asn Arg Arg Thr Lys Thr Lys Thr Leu Leu Ser Leu Leu Ser Phe Leu
      -20                -15                -10
gat gaa acc tct gga cta agc aca cat ctt cct tgt tta tct ctc tca      197
Asp Glu Thr Ser Gly Leu Ser Thr His Leu Pro Cys Leu Ser Leu Ser
      -5                1                5                10
aag gag tgt gga gtg ctt cat ctg gac atc cac ggg aag aag gaa gac      245
Lys Glu Cys Gly Val Leu His Leu Asp Ile His Gly Lys Lys Glu Asp
      15                20                25
atg aga gat gag gtc ttg ctg gcc ttg aac tyc tgc acc cac agg      290
Met Arg Asp Glu Val Leu Leu Ala Leu Asn Xaa Cys Thr His Arg
      30                35                40

```

<210> 86  
 <211> 336  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 100..336

<221> sig\_peptide  
 <222> 100..156  
 <223> Von Heijne matrix  
 score 9.899999961853027  
 seq ILFLVFLLAGLRS/KA

```

<400> 86
ccagatctgt tctgcaacat tcaccgttct ctgcatccag ctctgcttat ctgctgttac      60
cttggacacc agagcagcta taggtatctg ccagragcw atg aaa tca ttc agc      114
                                Met Lys Ser Phe Ser
                                -15
cgg atc ctc ttc ctc gtc ttc ctc ctc gcc ggc ctg agg tcc aag gcc      162
Arg Ile Leu Phe Leu Val Phe Leu Leu Ala Gly Leu Arg Ser Lys Ala
      -10                -5                1
gct ccc tca gcc cct ctg cct ttg ggc tgt ggc ttt ccg gac atg gcc      210
Ala Pro Ser Ala Pro Leu Pro Leu Gly Cys Gly Phe Pro Asp Met Ala
      5                10                15
cac ccc tct gag act tcc cct ctg aag ggt gct tct gaa aat tcc aaa      258
His Pro Ser Glu Thr Ser Pro Leu Lys Gly Ala Ser Glu Asn Ser Lys
      20                25                30
cga gat cgc ctt aac cca gaa ttt cct ggg act cct tac cct gag cct      306
Arg Asp Arg Leu Asn Pro Glu Phe Pro Gly Thr Pro Tyr Pro Glu Pro
      35                40                45                50
tcc aag cta cct cat acg gtt tcc ctg gaa      336
Ser Lys Leu Pro His Thr Val Ser Leu Glu
      55                60

```

<210> 87  
 <211> 262  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 108..260

<221> sig\_peptide

&lt;222&gt; 108..230

&lt;223&gt; Von Heijne matrix

score 9.89999961853027

seq SLCHLGWSAVVQS/QP

&lt;400&gt; 87

taggagtgga	gtgactgggt	gatatgataa	ctctatgttt	aacttttttaa	ggaactgcta	60
gacttttctg	aagtgactat	gccattttac	attaacacca	ggagtggt	atg agg gtg	116
					Met Arg Val	
					-40	
ccg att ttt	cca cat cct	cac caa ctc	tcg tta tta	ttc atc cat	tta	164
Pro Ile Phe	Pro His Pro	His Gln Leu	Ser Leu Leu	Phe Ile His	Leu	
	-35		-30		-25	
ttt att tat	tta ttt aga	gaa agg gtc	tct ctc tgt	cac cta ggc	tgg	212
Phe Ile Tyr	Leu Phe Arg	Glu Arg Val	Ser Leu Cys	His Leu Gly	Trp	
	-20		-15		-10	
agt gca gtg	gta caa tca	cag cca act	aca acc ttg	acc tcc cgc	gct	260
Ser Ala Val	Val Gln Ser	Gln Pro Thr	Thr Thr Thr	Leu Thr Ser	Arg Ala	
	-5		1		5	10
am						262

&lt;210&gt; 88

&lt;211&gt; 149

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 18..149

&lt;221&gt; sig\_peptide

&lt;222&gt; 18..128

&lt;223&gt; Von Heijne matrix

score 9.89999961853027

seq FLFVLFCFGGSRA/LL

&lt;400&gt; 88

ttcggagctt	gaccagc atg	tgg aag gag	agc tct cat	ggc tgc aat	aac	50
		Met Trp Lys	Glu Ser Ser	His Gly Cys	Asn Asn	
		-35		-30		
tta ggg agt	tcc tac ctg	gat gac act	ggg gta gga	agt ttt ctg	ttt	98
Leu Gly Ser	Ser Tyr Leu	Asp Thr Gly	Val Gly Ser	Phe Leu Phe		
	-25		-20		-15	
gtt ttg ttc	tgt ttc gga	ggg tcc cgt	gca ctt ctc	ttg cct gga	tct	146
Val Leu Phe	Cys Phe Gly	Gly Ser Arg	Ala Leu Leu	Leu Pro Gly	Ser	
	-10		-5		1	5
ggg						149
Gly						

&lt;210&gt; 89

&lt;211&gt; 315

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 236..313

&lt;221&gt; sig\_peptide

&lt;222&gt; 236..283

&lt;223&gt; Von Heijne matrix

score 9.69999980926514

seq FLCLLFYLIVSCG/AV

```

<400> 89
gtaaaagaca aataacttgt atggtttgca aaatgatctg aatatgtgct tttataacat      60
tcagaataca cccaaaagta aacttttaggt ttaatgtaca gtatgttttc tatgtaattg      120
ttttgaataa gtaaatamcat ybtacatggc ttaaaactga aaaacgtatt cctgttactt      180
cttgatgctt ttgagaaatg aataatgttt tctccctttt aaatggtagt acagc atg      238
                                         Met
cac act ttt ctg tgc ttg ctt ttt tat ctc ata gta tct tgt gga gct      286
His Thr Phe Leu Cys Leu Leu Phe Tyr Leu Ile Val Ser Cys Gly Ala
-15          -10          -5          1
gtt ttc tta aca gtc cct tct ccc caa gg      315
Val Phe Leu Thr Val Pro Ser Pro Gln
          5          10

<210> 90
<211> 179
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 24..179

<221> sig_peptide
<222> 24..140
<223> Von Heijne matrix
      score 9.69999980926514
      seq SIILXLXFPGLG/QA

<221> misc_feature
<222> 57
<223> n=a, g, c or t

<400> 90
agmrctctgg ggcagtctgc acc atg gcc tgg cac ccc act cct cct cct ctt      53
                                         Met Ala Trp His Pro Thr Pro Pro Pro Leu
                                         -35          -30
csb ncw cct cct cca ctg mac agg gwc tcy ctc cca gcc tgt gct gac      101
Xaa Xaa Pro Pro Pro Leu Xaa Arg Xaa Ser Leu Pro Ala Cys Ala Asp
          -25          -20          -15
tca atc atc ctc tgm ctc tgm ttc cct ggg atc ctc ggw caa gct cac      149
Ser Ile Ile Leu Xaa Leu Xaa Phe Pro Gly Ile Leu Gly Gln Ala His
          -10          -5          1
ctg mac tct gag cag tgg aca cag tac cta      179
Leu Xaa Ser Glu Gln Trp Thr Gln Tyr Leu
          5          10

<210> 91
<211> 423
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 311..421

<221> sig_peptide
<222> 311..373
<223> Von Heijne matrix
      score 9.69999980926514
      seq LHLILLSGTCFT/WI

```

60

```

<400> 91
gactctatag srcaaatggt taagaacata tacttgggag tcagttgatc tgggttcaaa      60
ttctagctgt gctactttct acctatgctg tattggacaa atgatactgt gtatctgttt      120
cttcaaccgt aagttgggta tattaatatc cttacctcaa aaggatcatga tgattaagtg      180
agtbaatgca tgtaaaatgc cttctgtgcc gggcagtcag aaaccactca ataaatattg      240
attattctca ccaaagatgt gcttcctgac ctcaaaagcc tgtcagccta atataaagac      300
agtggtgacaa atg cca atc ctg cct cag gac atc ttg cac ttg ctg atc      349
          Met Pro Ile Leu Pro Gln Asp Ile Leu His Leu Leu Ile
          -20          -15          -10
ctt ctg tct gga aca tgc ttc act tgg att ctt ttg tgg ctt cca ctc      397
Leu Leu Ser Gly Thr Cys Phe Thr Trp Ile Leu Leu Trp Leu Pro Leu
          -5          1          5
tcc cct ctg ttg ggc ctg aaa tgc ta      423
Ser Pro Leu Leu Gly Leu Lys Cys
          10          15

```

```

<210> 92
<211> 316
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> CDS
<222> 62..316
<221> sig_peptide
<222> 62..121
<223> Von Heijne matrix
      score 9.60000038146973
      seq LLALLLCGRPGRG/QT

```

```

<221> misc_feature
<222> 264,266
<223> n=a, g, c or t

```

```

<400> 92
accgcagctc cagagccctg cgggaggact cagagtcagg gacacagcag cgtccggcga      60
g atg aag gcg ctc ggg gct gtc ctg ctt gcb ctc ttg ctg tgc ggg cgg      109
      Met Lys Ala Leu Gly Ala Val Leu Leu Ala Leu Leu Leu Cys Gly Arg
      -20          -15          -10          -5
cca ggg aga ggg cag aca cag cag gag gaa gag gaa gag gac gag gac      157
Pro Gly Arg Gly Gln Thr Gln Gln Glu Glu Glu Glu Asp Glu Asp
          1          5          10
cac ggg cca gat gac tac gac gag gaa gat gag gat gag gtt gaa gag      205
His Gly Pro Asp Asp Tyr Asp Glu Glu Asp Glu Asp Glu Val Glu Glu
          15          20          25
gag gag acc aac agg ctc cct ggt ggc agg agc aga gtg ctg ctg cgg      253
Glu Glu Thr Asn Arg Leu Pro Gly Gly Arg Ser Arg Val Leu Leu Arg
          30          35          40
tgc tac acc tnk nag tcc ctg ccc agg gac gag cgc tgc aac ctg acg      301
Cys Tyr Thr Xaa Xaa Ser Leu Pro Arg Asp Glu Arg Cys Asn Leu Thr
          45          50          55          60
cag aac tgc tca cat      316
Gln Asn Cys Ser His
          65

```

```

<210> 93
<211> 508
<212> DNA
<213> Homo sapiens

```

```

<220>

```

<221> CDS  
<222> 245..508

<221> sig\_peptide  
<222> 245..289  
<223> Von Heijne matrix  
score 9.60000038146973  
seq EYVLLLFLALCSA/KP

<400> 93  
agtaactaaca tggactaatc tgtgggagca gtttattcca gtatcaccca ggggtgcagcc 60  
acaccaggac tgtgttgaag ggtgtttttt ttcttttaaa tgtaatacct cctcatcttt 120  
tcttcttaca cagtgtctga gaacatttac attatagata agtagtacat ggtggataac 180  
ttctactttt aggaggacta ctctcttctg acagtcctag actggctctc tacactaaga 240  
cacc atg aag gag tat gtg ctc cta tta ttc ctg gct ttg tgc tct gcc 289  
Met Lys Glu Tyr Val Leu Leu Leu Phe Leu Ala Leu Cys Ser Ala  
-15 -10 -5  
aaa ccc ttc ttt agc cct tca cac atc gca ctg aag aat atg atg ctg 337  
Lys Pro Phe Phe Ser Pro Ser His Ile Ala Leu Lys Asn Met Met Leu  
1 5 10 15  
aag gat atg gaa gac aca gat gat gat gat gat gat gat gat gat 385  
Lys Asp Met Glu Asp Thr Asp Asp Asp Asp Asp Asp Asp Asp Asp  
20 25 30  
gat gat gat gag gac aac tct ctt ttt cca aca aga gag cca aga agc 433  
Asp Asp Asp Glu Asp Asn Ser Leu Phe Pro Thr Arg Glu Pro Arg Ser  
35 40 45  
cat ttt ttt cca ttt gat ctg ttt cca atg tgt cca ttt gga tgt cag 481  
His Phe Phe Pro Phe Asp Leu Phe Pro Met Cys Pro Phe Gly Cys Gln  
50 55 60  
tgc tat tca cga gtt gta cat tgc tca 508  
Cys Tyr Ser Arg Val Val His Cys Ser  
65 70

<210> 94  
<211> 321  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 36..320

<221> sig\_peptide  
<222> 36..92  
<223> Von Heijne matrix  
score 9.60000038146973  
seq FLLLVAAPRWAMS/QV

<400> 94  
actttctgag aggccctggac ctccctgcaca agaac atg aaa cac ctg tgg ttc 53  
Met Lys His Leu Trp Phe  
-15  
ttc ctc ctc ctg gtg gca gct ccc aga tgg gcc atg tct cag gtg caa 101  
Phe Leu Leu Leu Val Ala Ala Pro Arg Trp Ala Met Ser Gln Val Gln  
-10 -5 1  
ctg cag gaa tcg ggc ccg aga ctg gtg aaa cct tcg ggg acc ctg tcc 149  
Leu Gln Glu Ser Gly Pro Arg Leu Val Lys Pro Ser Gly Thr Leu Ser  
5 10 15  
ctc acc tgc agt gtc tct ggt ggc tcc atg gcc act agt gac tgg tgg 197  
Leu Thr Cys Ser Val Ser Gly Gly Ser Met Ala Thr Ser Asp Trp Trp  
20 25 30 35  
agt tgg ttt cga cag acm ccg gag aag ggt ctg gag tgg att ggg gaa 245  
Ser Trp Phe Arg Gln Thr Pro Glu Lys Gly Leu Glu Trp Ile Gly Glu

[illegible]



&lt;223&gt; Von Heijne matrix

score 9.5

seq VLLFLILLYMSWS/AS

&lt;400&gt; 96

```

agagacacac ttggacgrtt cctgcagraa tcagtgaggc agtctcctcc caggggcttg      60
gsgcctggct cgaggcgagg ctgccggccc ggacgctgac tgcccagtgc cacagac      117
atg gcc aac ggg acc aac gcc tct gcc cca tac tac agc tat gaa tac      165
Met Ala Asn Gly Thr Asn Ala Ser Ala Pro Tyr Tyr Ser Tyr Glu Tyr
      -60      -55      -50
tac ctg gac tat ctg gac ctc att ccc gtg gac gag aag aag ctg aaa      213
Tyr Leu Asp Tyr Leu Asp Leu Ile Pro Val Asp Glu Lys Lys Leu Lys
      -45      -40      -35
gcc cac aaa cat tcc atc gtg atc gca ttc tgg gtg agc ctg gct gcc      261
Ala His Lys His Ser Ile Val Ile Ala Phe Trp Val Ser Leu Ala Ala
      -30      -25      -20
ttc gtg gtg ctg ctc ttc ctc atc ttg ctc tac atg tcc tgg tcc gcs      309
Phe Val Val Leu Leu Phe Leu Ile Leu Leu Tyr Met Ser Trp Ser Ala
      -15      -10      -5      1
tcc ccg      315
Ser Pro

```

&lt;210&gt; 97

&lt;211&gt; 460

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 62..460

&lt;221&gt; sig\_peptide

&lt;222&gt; 62..118

&lt;223&gt; Von Heijne matrix

score 9.39999961853027

seq FLFVVAATGVQS/QX

&lt;400&gt; 97

```

agcatcacat aacaaccasa ttctctctct aaagaagccc ctgggagcac agctcatcac      60
c atg gac tgg acc tgg agg ttc ctc ttt gtg gtg gca gca gct aca ggt      109
Met Asp Trp Thr Trp Arg Phe Leu Phe Val Val Ala Ala Ala Thr Gly
      -15      -10      -5
gtc cag tcm cag gks cas ctg gwg cag tct ggg gct gag gtg aag aag      157
Val Gln Ser Gln Xaa Xaa Leu Xaa Gln Ser Gly Ala Glu Val Lys Lys
      1      5      10
cct ggg tcc tcg gtg aaa gtc tcc tgc arg gcy tct gga ggc atc ytc      205
Pro Gly Ser Ser Val Lys Val Ser Cys Xaa Ala Ser Gly Gly Ile Xaa
      15      20      25
agc asc tat agc ttc aac tgg gtg cgm cag gcc cct gga cag ggg ttt      253
Ser Xaa Tyr Ser Phe Asn Trp Val Arg Gln Ala Pro Gly Gln Gly Phe
      30      35      40      45
gag tgg ttg gga agg atc atc ccc atc ctc ggt ata aca aac tac gca      301
Glu Trp Leu Gly Arg Ile Ile Pro Ile Leu Gly Ile Thr Asn Tyr Ala
      50      55      60
gag aag ttt cgg ggc aga ctc acg atc acc gtg gac aaa tcc acg cgt      349
Glu Lys Phe Arg Gly Arg Leu Thr Ile Thr Val Asp Lys Ser Thr Arg
      65      70      75
gtt gtt tac atg gag cag agc agt ctg aca tct gcg gac acg gcc gta      397
Val Val Tyr Met Glu Gln Ser Ser Leu Thr Ser Ala Asp Thr Ala Val
      80      85      90
tat tat tgt gcg aaa ccg act atg act tcg gaa cta cgg gtc tac tat      445
Tyr Tyr Cys Ala Lys Pro Thr Met Thr Ser Glu Leu Arg Val Tyr Tyr
      95      100      105

```

cag wct aca cta tgg  
Gln Xaa Thr Leu Trp  
110

460

<210> 98  
<211> 230  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 140..229

<221> sig\_peptide  
<222> 140..205  
<223> Von Heijne matrix  
score 9.39999961853027  
seq LLLLSAFTSQTVS/GQ

<400> 98  
aacagaacaa tatcaaataag ctaacttcac ccccaaccac agtccttgct gttggcattt 60  
actcaactag tctttaattc ctgttttgac aaactttata aggtgctaca agacagatga 120  
tttttcacca tctaccata atg tgg aac aga tat ttt gtc ttc tat ctc ctg 172  
Met Trp Asn Arg Tyr Phe Val Phe Tyr Leu Leu  
-20 -15

ctt ttg tca gcg ttt acg agt caa aca gta tcc gga caa aga aag aaa 220  
Leu Leu Ser Ala Phe Thr Ser Gln Thr Val Ser Gly Gln Arg Lys Lys  
-10 -5 1 5  
gga ccc cgg g 230  
Gly Pro Arg

<210> 99  
<211> 467  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 40..465

<221> sig\_peptide  
<222> 40..96  
<223> Von Heijne matrix  
score 9.39999961853027  
seq FLLLVAAPRWVLS/QL

<400> 99  
aaatactttc tgagagccct ggacctctg tgcaagaac atg aaa cac ctg ggg 54  
Met Lys His Leu Gly  
-15  
ttc ttc ctc ctg ctg gtg gca gct ccc aga tgg gtc ctg tcc cag ctg 102  
Phe Phe Leu Leu Leu Val Ala Ala Pro Arg Trp Val Leu Ser Gln Leu  
-10 -5 1  
cag ctc cag gag tcc ggc tca gga ctg gag aag cct tca cag acc ctg 150  
Gln Leu Gln Glu Ser Gly Ser Gly Leu Glu Lys Pro Ser Gln Thr Leu  
5 10 15  
tcc ctc acc tgc tct gtc tct ggt ggc tcc atc agt agt gat gat ttg 198  
Ser Leu Thr Cys Ser Val Ser Gly Gly Ser Ile Ser Ser Asp Asp Leu  
20 25 30  
tcg tgg agc tgg atc cga cag ccg cca ggg aag ggc ctg gag tgg att 246  
Ser Trp Ser Trp Ile Arg Gln Pro Pro Gly Lys Gly Leu Glu Trp Ile  
35 40 45 50  
ggc tac att tat caa aat gag agg acc ctc tac aac ccg tcc ctc aag 294

[illegible]

```
<210> 100
<211> 504
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 39..503
```

```
<221> sig_peptide
<222> 39..95
<223> Von Heijne matrix
      score 9.30000019073486
      seq FLLLVLVAGPRWVLS/QV
```

[illegible]

135

<210> 101  
 <211> 336  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 148..336

<221> sig\_peptide  
 <222> 148..270  
 <223> Von Heijne matrix  
 score 9.30000019073486  
 seq VLXLFCVFEEAES/RS

<400> 101  
 agagctcgcg gtggactccg acccggcgca acatggccgc agcctcgct ctgcgcgact 60  
 gccaggcctg gaaggatgag aggtcccgcc tctccaccac aagcaacgaa gcctgcaagc 120  
 tgttcgatgc cagctgacc cagggtat atg gcc tgc cga gag agg ccg cgg ccc 174  
 Met Ala Cys Arg Glu Arg Pro Arg Pro  
 -40 -35  
 ctt ctg tgg agg tct agg gga agg ttt ttt aat tgg gga aag ctg ttt 222  
 Leu Leu Trp Arg Ser Arg Gly Arg Phe Phe Asn Trp Gly Lys Leu Phe  
 -30 -25 -20  
 ttt tgt ttt gtt ttg mtt ttg ttt tgt ttt gtt ttt gag gcg gag tct 270  
 Phe Cys Phe Val Leu Xaa Leu Phe Cys Phe Val Phe Glu Ala Glu Ser  
 -15 -10 -5  
 cgc tct gtc gcc cag gct gga gtg cag tgg cgc tat ttc ggc tca cta 318  
 Arg Ser Val Ala Gln Ala Gly Val Gln Trp Arg Tyr Phe Gly Ser Leu  
 1 5 10 15  
 caa gct ttg cct ccc tgg 336  
 Gln Ala Leu Pro Pro Trp  
 20

<210> 102  
 <211> 289  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 214..288

<221> sig\_peptide  
 <222> 214..276  
 <223> Von Heijne matrix  
 score 9.19999980926514  
 seq FILFHWSLCLCLC/QY

<400> 102  
 cccttatgtt ttcttttagt aatttcatag tttcaagttt tatatataag tctttaatcc 60  
 attttgagtt gatttggtga tatgggtggag acagggtcta gtcttggtct tctgcatgtg 120  
 actttccaat tttccagca ccatttattg gagaaactgt ctttttccca gtgcatgttc 180  
 ttggcacctt tgttgaaaaa cagttggcca tag atg cat gaa ttt att tct ggg 234  
 Met His Glu Phe Ile Ser Gly  
 -20 -15  
 ttc ttt att ctc ttt cat tgg tct ctg tgt ttg tgt tta tgc caa tac 282  
 Phe Phe Ile Leu Phe His Trp Ser Leu Cys Leu Cys Leu Cys Gln Tyr  
 -10 -5 1  
 cat gcc g 289  
 His Ala

<210> 103  
 <211> 383  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 252..383

<221> sig\_peptide  
 <222> 252..377  
 <223> Von Heijne matrix  
 score 9.19999980926514  
 seq LLVCLFAVTSILC/SS

<400> 103  
 atcctccagc taataagtgt ccaagctggg actcaaactt gggcctttta actgtgctgc 60  
 tattctacct ctcccttgct ctttccagac caggcttggg acataacact aacacccttt 120  
 tcctttcatt tcattctctg tccttcagtc attcctaaac attgacaybc attgagttcc 180  
 ttggctctgg ccatagtcct ttctcccttt cccctctggg gcatcaaata gtgattacag 240  
 tatccacagg g atg gca tat gcc att tca cca ttt cac agt tcc tgg aat 290  
 Met Ala Tyr Ala Ile Ser Pro Phe His Ser Ser Trp Asn  
 -40 -35 -30  
 cca ctt ttc act tct cat aaa gct tca gca agc cat tct cat ctt ggg 338  
 Pro Leu Phe Thr Ser His Lys Ala Ser Ala Ser His Ser His Leu Gly  
 -25 -20 -15  
 ttg ctt gtt tgc cta ttt gct gtt aca tcc att ctc tgc tcc tca 383  
 Leu Leu Val Cys Leu Phe Ala Val Thr Ser Ile Leu Cys Ser Ser  
 -10 -5 1

<210> 104  
 <211> 211  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 30..209

<221> sig\_peptide  
 <222> 30..74  
 <223> Von Heijne matrix  
 score 9.19999980926514  
 seq PVLLLALLGFILP/LP

<221> misc\_feature  
 <222> 83  
 <223> n=a, g, c or t

<400> 104  
 agaaagagat taccagccac agacgggtc atg agc ccg gta tta ctg ctg gcc 53  
 Met Ser Pro Val Leu Leu Ala  
 -15 -10  
 ctc ctg ggg ttc atc ctc cca ctg cca ggn agt gca rgc gct gss tck 101  
 Leu Leu Gly Phe Ile Leu Pro Leu Pro Gly Ser Ala Xaa Ala Xaa Ser  
 -5 1 5  
 gcc agt ttg gga cag ttc agc atg tgt gga agg tgt ccg acm tgc ccc 149  
 Ala Ser Leu Gly Gln Phe Ser Met Cys Gly Arg Cys Pro Thr Cys Pro  
 10 15 20 25  
 ggc aat gga ccc cta aga aca cca gct gcg aca sgg vtt rgg gtg cca 197  
 Gly Asn Gly Pro Leu Arg Thr Pro Ala Ala Thr Xaa Xaa Xaa Val Pro

68

30 35 40 211  
 gga cac gtt gat gc  
 Gly His Val Asp  
 45

<210> 105  
 <211> 492  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 29..490

<221> sig\_peptide  
 <222> 29..97  
 <223> Von Heijne matrix  
 score 9.10000038146973  
 seq SLLLFSLMCETSA/FY

<400> 105  
 agtcattacg gcgacacgtg gatccaag atg gcg acg gcg atg gat tgg ttg 52  
 Met Ala Thr Ala Met Asp Trp Leu  
 -20  
 ccg tgg tct tta ctg ctt ttc tcc ctg atg tgt gaa aca agc gcc ttc 100  
 Pro Trp Ser Leu Leu Leu Phe Ser Leu Met Cys Glu Thr Ser Ala Phe  
 -15 -10 -5 1  
 tat gtg cct ggg gtc gcg cct atc aac ttc cac cag aac gat ccc gta 148  
 Tyr Val Pro Gly Val Ala Pro Ile Asn Phe His Gln Asn Asp Pro Val  
 5 10 15  
 gaa atc aag gct gtg aag ctc acc agc tct cga acc cag cta cct tat 196  
 Glu Ile Lys Ala Val Lys Leu Thr Ser Ser Arg Thr Gln Leu Pro Tyr  
 20 25 30  
 gaa tac tat tca ctg ccc ttc tgc cag ccc agc aag ata acc tac aag 244  
 Glu Tyr Thr Ser Leu Pro Phe Cys Gln Pro Ser Lys Ile Thr Tyr Lys  
 35 40 45  
 gca gag aat ctg gga gag gtg ctg aga ggg gac cgg att gtc aac acc 292  
 Ala Glu Asn Leu Gly Glu Val Leu Arg Gly Asp Arg Ile Val Asn Thr  
 50 55 60 65  
 cct ttc cag gtt ctc atg aac agc gag aag aag tgt gaa gtt ctg tgc 340  
 Pro Phe Gln Val Leu Met Asn Ser Glu Lys Lys Cys Glu Val Leu Cys  
 70 75 80  
 agc cag tcc aac aag cca gtg acc ctg aca gtg gag cag agc cga ctc 388  
 Ser Gln Ser Asn Lys Pro Val Thr Leu Thr Val Glu Gln Ser Arg Leu  
 85 90 95  
 gtg gcc gag cgg atc aca gaa gac tac tac gtc cac ctc att gct gac 436  
 Val Ala Glu Arg Ile Thr Glu Asp Tyr Tyr Val His Leu Ile Ala Asp  
 100 105 110  
 aac ctg cct gtg gcc acc ggc tgg agc tct act cca acc gag aca gcg 484  
 Asn Leu Pro Val Ala Thr Gly Trp Ser Ser Thr Pro Thr Glu Thr Ala  
 115 120 125  
 atg aca ag 492  
 Met Thr  
 130

<210> 106  
 <211> 126  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 41..124

```

<221> sig_peptide
<222> 41..94
<223> Von Heijne matrix
      score 9.10000038146973
      seq LISLLQCAHVSLG/LQ

<400> 106
tccatttgca gcccatgggt gtcacatcac gctgtttata atg ccc agc ccc tgc      55
                                         Met Pro Ser Pro Cys
                                         -15
ctg atc tct ctt ctt caa tgt gct cat gtg tcc ctt ggc tta cag tat      103
Leu Ile Ser Leu Leu Gln Cys Ala His Val Ser Leu Gly Leu Gln Tyr
      -10                               -5                               1
cca tgc stt ctc ctt ctc cct cc      126
Pro Cys Xaa Leu Leu Leu Pro
      5                               10

<210> 107
<211> 242
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 82..240

<221> sig_peptide
<222> 82..132
<223> Von Heijne matrix
      score 9.10000038146973
      seq LVLAAPCLGIASA/VP

<400> 107
accagaccgc ggacgtctgt aatctcagag gcttgtttgc tgagggtgcc tgcgcastgc      60
gacggctgct ggttttgaaa c atg aat ctt tcg ctc gtc ctg gct gcc ttt      111
                        Met Asn Leu Ser Leu Val Leu Ala Ala Phe
                        -15                               -10
tgc ttg gga ata gcc tcc gct gtt cca aaa ttt gac caa aat ttg gat      159
Cys Leu Gly Ile Ala Ser Ala Val Pro Lys Phe Asp Gln Asn Leu Asp
      -5                               1                               5
aca aag tgg tac cag tgg aag gca aca cac aga aga tta tat ggc gcg      207
Thr Lys Trp Tyr Gln Trp Lys Ala Thr His Arg Arg Leu Tyr Gly Ala
      10                               15                               20                               25
aat gaa gaa gga tgg agg aga gca gcg tgg gag gg      242
Asn Glu Glu Gly Trp Arg Arg Ala Ala Trp Glu
      30                               35

<210> 108
<211> 336
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 81..335

<221> sig_peptide
<222> 81..137
<223> Von Heijne matrix
      score 9
      seq WVFLVAIFTGVHC/EV

```

```

<400> 108
agctctggga gaggagcccc agccgtgaga ttcccagaag tttccacttg gtgatcagca      60
ctgaacacag accaccaacc atg gag ttt ggc ctt aat tgg gtt ttc ctt gtt      113
                               Met Glu Phe Gly Leu Asn Trp Val Phe Leu Val
                               -15                               -10

gct att ttt aca ggt gtc cac tgt gag gtg cag ttg gtg gag tct ggg      161
Ala Ile Phe Thr Gly Val His Cys Glu Val Gln Leu Val Glu Ser Gly
                               -5                               1                               5

gga gac ctg gta cag cca ggg cgg tcc ctg aga ctc tcc tgt aca gct      209
Gly Asp Leu Val Gln Pro Gly Arg Ser Leu Arg Leu Ser Cys Thr Ala
                               10                               15                               20

tct gga ttc acc ttt ggt gat tat gcc atg acc tgg ttc cgc cag gct      257
Ser Gly Phe Thr Phe Gly Asp Tyr Ala Met Thr Trp Phe Arg Gln Ala
25                               30                               35                               40

tca ggg aag cga ctg gag tgg cta ggt ttc att aga aat aga ggt tcs      305
Ser Gly Lys Arg Leu Glu Trp Leu Gly Phe Ile Arg Asn Arg Gly Ser
                               45                               50                               55

ggt ggg tca gca gag tac ggc gcg tct gtg a      336
Gly Gly Ser Ala Glu Tyr Gly Ala Ser Val
                               60                               65

```

```

<210> 109
<211> 160
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> CDS
<222> 6..158

<221> sig_peptide
<222> 6..56
<223> Von Heijne matrix
score 9
seq LLILLMLLLFAIH/IN

```

```

<400> 109
cagct atg aaa aac tgc cta ctc ata ctc ctc atg ctt ctc tta ttt gca      50
Met Lys Asn Cys Leu Leu Ile Leu Leu Met Leu Leu Leu Phe Ala
                               -15                               -10                               -5

ata cac ata aac cgt atg aat gta agg aat gtg gga aat act tta gtc      98
Ile His Ile Asn Arg Met Asn Val Arg Asn Val Gly Asn Thr Leu Val
1                               5                               10

gta gtg caa atc tta ttc agc atc aga gta ttc ata ctg gag aga aac      146
Val Val Gln Ile Leu Phe Ser Ile Arg Val Phe Ile Leu Glu Arg Asn
15                               20                               25                               30

cct ttg aat gtg gg      160
Pro Leu Asn Val

```

```

<210> 110
<211> 527
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> CDS
<222> 81..527

<221> sig_peptide
<222> 81..137
<223> Von Heijne matrix
score 9
seq WIFLLAILKGVQC/EV

```



<221> misc\_feature  
 <222> 307..308,466..467  
 <223> n=a, g, c or t

<400> 110  
 agctctggga gaggagcccc agccctgaga ttcccaggtg tttccattca gtgatcagca 60  
 ctgaacacag aggactcacc atg gag ttg gga ctg agc tgg att ttc ctt ttg 113  
 Met Glu Leu Gly Leu Ser Trp Ile Phe Leu Leu  
 -15 -10  
 gct att tta aaa ggt gtc cag tgt gaa gtg cag ctg gtg gag tct ggg 161  
 Ala Ile Leu Lys Gly Val Gln Cys Glu Val Gln Leu Val Glu Ser Gly  
 -5 1 5  
 gga ggc ttg gta cag cct ggc agg tcc ctg aga ctc tcc tgt gca gcc 209  
 Gly Gly Leu Val Gln Pro Gly Arg Ser Leu Arg Leu Ser Cys Ala Ala  
 10 15 20  
 tct gga ttc acc ttt gat gat tac gcc atg cac tgg gtc cgg caa gct 257  
 Ser Gly Phe Thr Phe Asp Asp Tyr Ala Met His Trp Val Arg Gln Ala  
 25 30 35 40  
 cca ggg aag ggc ctg gag tgg gtc tca gga att act tgg aat agt ggt 305  
 Pro Gly Lys Gly Leu Glu Trp Val Ser Gly Ile Thr Trp Asn Ser Gly  
 45 50 55  
 ann ata ggc tac gcg gac tct gtg aag ggc cga ttc acc atc tcc aga 353  
 Xaa Ile Gly Tyr Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg  
 60 65 70  
 gac aac gcc aag aac tcc ctg tat ttg caa atg aac agt ctg aga act 401  
 Asp Asn Ala Lys Asn Ser Leu Tyr Leu Gln Met Asn Ser Leu Arg Thr  
 75 80 85  
 gag gac acg gcc ttc tat ttc tgt gca aaa gct cgc ggg ctc ttt agc 449  
 Glu Asp Thr Ala Phe Tyr Phe Cys Ala Lys Ala Arg Gly Leu Phe Ser  
 90 95 100  
 gat acc tgg ccc tac vnn cac tac gct atg gac gtc tgg ggc caa ggg 497  
 Asp Thr Trp Pro Tyr Xaa His Tyr Ala Met Asp Val Trp Gly Gln Gly  
 105 110 115 120  
 acc acg gtc acc gtc tcc tca gcc tcc acc 527  
 Thr Thr Val Thr Val Ser Ser Ala Ser Thr  
 125 130

<210> 111  
 <211> 154  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 80..154

<221> sig\_peptide  
 <222> 80..121  
 <223> Von Heijne matrix  
 score 8.89999961853027  
 seq LLVFFVLWTCSLA/LL

<400> 111  
 ctggaaaggg aggagccaaa aggggaacgc tttcttgatt gtcccagcct cattaggagc 60  
 taccacaggg ctctcctgc atg ctc ctt gtt ttc ttt gtg ctc tgg act tgc 112  
 Met Leu Leu Val Phe Phe Val Leu Trp Thr Cys  
 -10 -5  
 tca ctt gca ctg ctt gct tct tcc cca atc gcm gcc yac cca 154  
 Ser Leu Ala Leu Leu Ala Ser Ser Pro Ile Ala Ala Xaa Pro  
 1 5 10

<210> 112  
 <211> 441  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 59..439

<221> sig\_peptide  
 <222> 59..115  
 <223> Von Heijne matrix  
 score 8.89999961853027  
 seq ILLLVAAATDASS/QM

<400> 112  
 atcacccatc aaccacatcc ctctctctaga gagtcccctg aaagcacagc tcctcacc 58  
 atg gac tgg acc tgg aga atc ctc ctc ttg gtg gca gca gcc aca gat 106  
 Met Asp Trp Thr Trp Arg Ile Leu Leu Val Ala Ala Ala Thr Asp  
 -15 -10 -5  
 gcc tcc tcc cag atg cag ctg ttg cag tct ggg cct gaa gtg aag aag 154  
 Ala Ser Ser Gln Met Gln Leu Leu Gln Ser Gly Pro Glu Val Lys Lys  
 1 5 10  
 act ggg tcc tca gtg aaa ctt tcc tgc acg gcc tcc ggc gac acc ctc 202  
 Thr Gly Ser Ser Val Lys Leu Ser Cys Thr Ala Ser Gly Asp Thr Leu  
 15 20 25  
 gcc tac cac tac ctg cac tgg gtg cga cag gcc ccc gga caa gcg ctt 250  
 Ala Tyr His Tyr Leu His Trp Val Arg Gln Ala Pro Gly Gln Ala Leu  
 30 35 40 45  
 gag tgg atg gga tgg atc aca cct ttc agt gga gac acc aac ttc gca 298  
 Glu Trp Met Gly Trp Ile Thr Pro Phe Ser Gly Asp Thr Asn Phe Ala  
 50 55 60  
 cag cga ttc cag gac aga ctc acc ttc acc agg gac agg tct atg agc 346  
 Gln Arg Phe Gln Asp Arg Leu Thr Phe Thr Arg Asp Arg Ser Met Ser  
 65 70 75  
 aca gtc tac atg acc ctg acc agc ctg ata tct gaa gac aca gcc atg 394  
 Thr Val Tyr Met Thr Leu Thr Ser Leu Ile Ser Glu Asp Thr Ala Met  
 80 85 90  
 tat tac tgt gcc act gat gga cgt cgc acc aac cgt ctt ttt gaa ca 441  
 Tyr Tyr Cys Ala Thr Asp Gly Arg Arg Thr Asn Arg Leu Phe Glu  
 95 100 105

<210> 113  
 <211> 369  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 164..367

<221> sig\_peptide  
 <222> 164..217  
 <223> Von Heijne matrix  
 score 8.80000019073486  
 seq LGCLLWLLTHIKA/QD

<221> misc\_feature  
 <222> 290..292  
 <223> n=a, g, c or t

<400> 113

73

```

cagtttcagt ttctctccct tcctagtaga gacaaaaagg agacacattt tatccgtgca      60
tccaaagact ccgatgttgg tcatggactt gggaagacag tcttcccttg gcgtttgatc      120
actgcggaga tgccttcctt gatcattcac ccacattccc ttg atg gca ggt caa      175
                               Met Ala Gly Gln
                               -15
ttg ctg gga tgc ctg ctt tgg ctg ctc acc cac att aaa gcc cag gac      223
Leu Leu Gly Cys Leu Leu Trp Leu Leu Thr His Ile Lys Ala Gln Asp
                               -10                               -5                               1
tca gtc agg gat gcc tac tgg aag act ggt agc tgc cca cct cca ttt      271
Ser Val Arg Asp Ala Tyr Trp Lys Thr Gly Ser Cys Pro Pro Pro Phe
                               5                               10                               15
ctc cat gtg tct acc ttc nnn kkt aaa ctt acc ttc tcc act aag ggc      319
Leu His Val Ser Thr Phe Xaa Lys Leu Thr Phe Ser Thr Lys Gly
                               20                               25                               30
aac ctt ctg cat tcc att cct ctc tct tcc ccc tta gcc tgt gtt ctt      367
Asn Leu Leu His Ser Ile Pro Leu Ser Ser Pro Leu Ala Cys Val Leu
35                               40                               45                               50
ag                                                                 369

```

&lt;210&gt; 114

&lt;211&gt; 334

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 20..334

&lt;221&gt; sig\_peptide

&lt;222&gt; 20..292

&lt;223&gt; Von Heijne matrix

score 8.80000019073486

seq LFLMLLGELGVFA/SY

&lt;221&gt; misc\_feature

&lt;222&gt; 295

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 114

```

agctctgaat tgggaaggg atg aag gag gct gtg cct ccg ggt tgc acg aag      52
                               Met Lys Glu Ala Val Pro Pro Gly Cys Thr Lys
                               -90                               -85
agt ccg agt cat ttc tca gaa ggt ttt gat agg tgg gcc tta gag gag      100
Ser Pro Ser His Phe Ser Glu Gly Phe Asp Arg Trp Ala Leu Glu Glu
-80                               -75                               -70                               -65
acg ccg ccg gaa aac ctg att ggc gcc ctc ttg gcg atc ttc ggg cac      148
Thr Pro Pro Glu Asn Leu Ile Gly Ala Leu Leu Ala Ile Phe Gly His
                               -60                               -55                               -50
ctc gtg gtc agc att gca ctt aac ctc cag aag tac tgc cac atc cgc      196
Leu Val Val Ser Ile Ala Leu Asn Leu Gln Lys Tyr Cys His Ile Arg
                               -45                               -40                               -35
ctg gca ggc tcc aag gat ccc cgg gcc tat ttc aag acc aag aca tgg      244
Leu Ala Gly Ser Lys Asp Pro Arg Ala Tyr Phe Lys Thr Lys Thr Trp
                               -30                               -25                               -20
tgg ctg ggc ctg ttc ctg atg ctt ctg ggc gag ctg ggt gtg ttc gcm      292
Trp Leu Gly Leu Phe Leu Met Leu Leu Gly Glu Leu Gly Val Phe Ala
-15                               -10                               -5
tcn tac gcc ttc gcg ccg ctg tca ctc atc gtg ccc ctc agc      334
Ser Tyr Ala Phe Ala Pro Leu Ser Leu Ile Val Pro Leu Ser
1                               5                               10

```

&lt;210&gt; 115

<211> 153  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 21..152

<221> sig\_peptide  
 <222> 21..74  
 <223> Von Heijne matrix  
 score 8.80000019073486  
 seq LLSCWALLGTTFG/CG

<400> 115  
 acaccctgc cagcggcacc atg gct ttc ctc tgg ctc ctc tcc tgc tgg gcc 53  
 Met Ala Phe Leu Trp Leu Leu Ser Cys Trp Ala  
 -15 -10  
 ctc ctg ggt acc acc ttc ggc tgc ggg gtc ccc gcc atc cac cct ggc 101  
 Leu Leu Gly Thr Thr Phe Gly Cys Gly Val Pro Ala Ile His Pro Gly  
 -5 1 5  
 tgc caa ctg agc ccg cgg ctc cct ccg acc ctg ctc ccc aca gag cgc 149  
 Cys Gln Leu Ser Pro Arg Leu Pro Pro Thr Leu Leu Pro Thr Glu Arg  
 10 15 20 25  
 ggg g 153  
 Gly

<210> 116  
 <211> 292  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 47..292

<221> sig\_peptide  
 <222> 47..106  
 <223> Von Heijne matrix  
 score 8.80000019073486  
 seq LWLFFVLNLGSFA/FS

<400> 116  
 taccagtaac ttctttcatg gttcaataaa atcatagctt tagttt atg gca cct 55  
 Met Ala Pro  
 -20  
 ttt caa aac ttc ctt tgg ctt ttc ttt gtg ctt aat tta ggw agt ttt 103  
 Phe Gln Asn Phe Leu Trp Leu Phe Phe Val Leu Asn Leu Gly Ser Phe  
 -15 -10 -5  
 gct ttt agc tca mtt ccd aat tct ctt ttt tac aca att cat ttt ggt 151  
 Ala Phe Ser Ser Xaa Pro Asn Ser Leu Phe Tyr Thr Ile His Phe Gly  
 1 5 10 15  
 cct aat ttc ttt act tta cta tat aaa caa ggt gct gaa atg tgt gtg 199  
 Pro Asn Phe Phe Thr Leu Leu Tyr Lys Gln Gly Ala Glu Met Cys Val  
 20 25 30  
 tat gta ttt aac ttc ctc tac cca ttt gct ctt ggt tat ttc ttc agt 247  
 Tyr Val Phe Asn Phe Leu Tyr Pro Phe Ala Leu Gly Tyr Phe Phe Ser  
 35 40 45  
 tat gat att ctg gat ttg cca gtc akt gtc cgt cct cct agc ggg 292  
 Tyr Asp Ile Leu Asp Leu Pro Val Xaa Val Arg Pro Pro Ser Gly  
 50 55 60

<210> 117

<211> 304  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 141..302

<221> sig\_peptide  
 <222> 141..245  
 <223> Von Heijne matrix  
 score 8.69999980926514  
 seq LLLSVAFNQLVFA/LY

<400> 117  
 tttctcatca atttcttgct tctctggcaa cctcaacctc tgattcctga ggccaataaa 60  
 actgaaactt tetgcttgag ctcttgtttt gccaggctga tggggctgag gtgcaccctc 120  
 tgaggaaaag ctgtaaatac atg gat ttt acc caa tgc cat tcc ctt ctt tta 173  
 Met Asp Phe Thr Gln Cys His Ser Leu Leu Leu  
 -35 -30 -25  
 agg gtt gaa tat tct cca gtg tct gtc tgc ttt tta tta ctt tcc gtt 221  
 Arg Val Glu Tyr Ser Pro Val Ser Val Cys Phe Leu Leu Leu Ser Val  
 -20 -15 -10  
 gcc ttc aat cag ttg gtt ttt gct ttg tat cca ata caa gct acw btc 269  
 Ala Phe Asn Gln Leu Val Phe Ala Leu Tyr Pro Ile Gln Ala Thr Xaa  
 -5 1 5  
 tgt ttc tct dda gtt tct ctc cct ttc ccc gct ca 304  
 Cys Phe Ser Xaa Val Ser Leu Pro Phe Pro Ala  
 10 15

<210> 118  
 <211> 145  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 76..144

<221> sig\_peptide  
 <222> 76..120  
 <223> Von Heijne matrix  
 score 8.5  
 seq LLLLACGVPSLWP/FA

<400> 118  
 gtgaaggtag gagtgttggg gccctgaccc ccgcaggagg gatgggcgga ttcgaggact 60  
 ggctgcctgc ccatac atg ctc ttg ctc ctg ctg gcc tgt ggt gtt ccc agc 111  
 Met Leu Leu Leu Leu Leu Ala Cys Gly Val Pro Ser  
 -15 -10 -5  
 ctg tgg ccc ttt gcw ctt gct ctc tta aag acc c 145  
 Leu Trp Pro Phe Ala Leu Ala Leu Leu Lys Thr  
 1 5

<210> 119  
 <211> 288  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 160..288

```

<221> sig_peptide
<222> 160..228
<223> Von Heijne matrix
      score 8.5
      seq SFLLLHLCQVLLS/RR

<400> 119
tcttttagtcc tgtttatatg atgaatcaca tttattgatt tgcatatggt gaaccatcct      60
tgtatcccag ggataaagcc tacttgattg taatggataa gcttcatgat gtgctgctga      120
atttggtttg ccagtatttt gttaaggatt tttacatca atg ttc att gag aat      174
                                   Met Phe Ile Glu Asn
                                   -20
att ggv ctg aag ttt tct ttt ttg ttg ttg cat ctc tgc cag gtt ttg      222
Ile Gly Leu Lys Phe Ser Phe Leu Leu Leu His Leu Cys Gln Val Leu
      -15      -10      -5
cta tca aga cga gct ggt acc att cct act gaa aca att cca aaa aaa      270
Leu Ser Arg Arg Ala Gly Thr Ile Pro Thr Glu Thr Ile Pro Lys Lys
      1      5      10
ttg agg agg aga gac ggg      288
Leu Arg Arg Arg Asp Gly
15      20

<210> 120
<211> 386
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 71..385

<221> sig_peptide
<222> 71..142
<223> Von Heijne matrix
      score 8.5
      seq XALLMGFLMVCLG/AF

<400> 120
aattctcttg gagcaagcag ggaagcagag gagcagcagg gtcagggtgc tgggttccta      60
aggtgcaagg atg cag aac aga act ggc ctc att ctc tgt gct ytt gcc      109
      Met Gln Asn Arg Thr Gly Leu Ile Leu Cys Ala Xaa Ala
      -20      -15
ctc ctg atg ggt ttc ctg atg gtc tgc ctg ggg gcc ttc ttc att tcc      157
Leu Leu Met Gly Phe Leu Met Val Cys Leu Gly Ala Phe Phe Ile Ser
      -10      -5      1      5
tgg ggc tcc ata ttc gac tgt cag ggg agc ctg att gcg gcc tat ttg      205
Trp Gly Ser Ile Phe Asp Cys Gln Gly Ser Leu Ile Ala Ala Tyr Leu
      10      15      20
ctt ctg cct ctg ggg ttt gtg atc ctt ctg agt gga att ttc tgg agc      253
Leu Leu Pro Leu Gly Phe Val Ile Leu Leu Ser Gly Ile Phe Trp Ser
      25      30      35
aac tat cgc cag gtg act gaa agc aaa gga gtg ttg agg cac atg ctc      301
Asn Tyr Arg Gln Val Thr Glu Ser Lys Gly Val Leu Arg His Met Leu
      40      45      50
cga caa cac ctt gct cat ggg gcc ctg ccc gtg gcc aca gta gac agt      349
Arg Gln His Leu Ala His Gly Ala Leu Pro Val Ala Thr Val Asp Ser
      55      60      65
gct gct ctt ctg aaa atc atg tgt aag car ttg ctt t      386
Ala Ala Leu Leu Lys Ile Met Cys Lys Gln Leu Leu
70      75      80

<210> 121
<211> 190

```

<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 34..189

<221> sig\_peptide  
<222> 34..165  
<223> Von Heijne matrix  
score 8.5  
seq LTCTSSLLSFALG/RS

<400> 121  
atcttgaaaa cggaaaataa aaacagcaga cct atg aag gtc gaa ggg gaa gaa 54  
Met Lys Val Glu Gly Glu Glu  
-40  
aag ctg tat cga ttg ttg aga tct ggc gac ttg ttt aaa ttt cat cag 102  
Lys Leu Tyr Arg Leu Leu Arg Ser Gly Asp Leu Phe Lys Phe His Gln  
-35 -30 -25  
cct cac ttc tat gaa ctc tca ggc ctc acg tgt acc agc tct ctg ctc 150  
Pro His Phe Tyr Glu Leu Ser Gly Leu Thr Cys Thr Ser Ser Leu Leu  
-20 -15 -10  
tcc ttt gcc ttg gga cgt tcc atc cct gga agt ttc cca g 190  
Ser Phe Ala Leu Gly Arg Ser Ile Pro Gly Ser Phe Pro  
-5 1 5

<210> 122  
<211> 211  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 32..211

<221> sig\_peptide  
<222> 32..88  
<223> Von Heijne matrix  
score 8.5  
seq LLLFSGAVALIQT/WA

<400> 122  
agattctccc cagacgccaa ggttgccgggt c atg gag tcc cga acc ctc ctc 52  
Met Glu Ser Arg Thr Leu Leu  
-15  
ctg ctg ttc tcg gga gcc gtg gcc ctg atc cag acc tgg gca ggt gag 100  
Leu Leu Phe Ser Gly Ala Val Ala Leu Ile Gln Thr Trp Ala Gly Glu  
-10 -5 1  
tgc ggg gtc ggg agg gaa aag gcc tct gcg gga agg agc gag ggg ccc 148  
Cys Gly Val Gly Arg Glu Lys Ala Ser Ala Gly Arg Ser Glu Gly Pro  
5 10 15 20  
gcc cgg agg agt aaa tct gca cat ata kbt aat tac aga tta caa tta 196  
Ala Arg Arg Ser Lys Ser Ala His Ile Xaa Asn Tyr Arg Leu Gln Leu  
25 30 35  
caa tca agg cag ggg 211  
Gln Ser Arg Gln Gly  
40

<210> 123  
<211> 353  
<212> DNA  
<213> Homo sapiens

<220>  
 <221> CDS  
 <222> 249..353

<221> sig\_peptide  
 <222> 249..296  
 <223> Von Heijne matrix  
 score 8.39999961853027  
 seq SVPLLCFWSLCYC/FA

<221> misc\_feature  
 <222> 187  
 <223> n=a, g, c or t

<400> 123  
 agcgagtcct tgcctcccgg cggctcagga cgagggcaga tctcgttctg gggcaagccg 60  
 ttgacactcg ctccctgcc a cgcgccgggc tccgtgccgc caagttttca tttccacct 120  
 tctctgcctc cagtccccca gcccttgccc gagagaaggg tcttaccggc cgggattgct 180  
 ggaaacnaa gaggtggttt ttgtttttta aaacttctgt ttcttgggag ggggtgtggc 240  
 ggggcagg atg agc aac tcc gtt cct ctg ctc tgt ttc tgg agc ctc tgc 290  
 Met Ser Asn Ser Val Pro Leu Leu Cys Phe Trp Ser Leu Cys  
 -15 -10 -5  
 tat tgc ttt gct gcg ggg agc ccc gta cct ttt ggt cca gag gga cgg 338  
 Tyr Cys Phe Ala Ala Gly Ser Pro Val Pro Phe Gly Pro Glu Gly Arg  
 1 5 10  
 ctg gaa gat aag ctc 353  
 Leu Glu Asp Lys Leu  
 15

<210> 124  
 <211> 249  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 93..248

<221> sig\_peptide  
 <222> 93..134  
 <223> Von Heijne matrix  
 score 8.39999961853027  
 seq PWTILLFAAGSLA/IP

<400> 124  
 tttttccccc atctggcctc acaggaggag ttggcgggga gccttgggcc cctctggcct 60  
 cagccggatt tcccagccaa acgcagagag ag atg ccc tgg acc atc ttg ctc 113  
 Met Pro Trp Thr Ile Leu Leu  
 -10  
 ttt gca gct ggc tcc ttg gcg atc cca gca cca tcc atc cgg gtg gtg 161  
 Phe Ala Ala Gly Ser Leu Ala Ile Pro Ala Pro Ser Ile Arg Val Val  
 -5 1 5  
 ccc ccg tac cca agc agc caa gag gac ccc atc cac atc gca tgc atg 209  
 Pro Pro Tyr Pro Ser Ser Gln Glu Asp Pro Ile His Ile Ala Cys Met  
 10 15 20 25  
 gcc gct ggg aac ttc ccg ggg gcg aat ttc aca ctg tat c 249  
 Ala Ala Gly Asn Phe Pro Gly Ala Asn Phe Thr Leu Tyr  
 30 35

<210> 125  
 <211> 375



<212> DNA  
<213> Homo sapiens

```
<220>  
<221> CDS  
<222> 175..375
```

```
<221> sig_peptide
<222> 175..366
<223> Von Heijne matrix
      score 8.39999961853027
      seq GFLFFGFLFPVFS/FP
```

<400>	125	
gtgctgcggc attcacgtga tctgcacggg cgcagatgta ggcaccggtc cgagtgcctg		60
ccctctgtcc ccgcggctgg gtctcgctctg ctccggttcc tgggctccta attcttggtc		120
cagcttcttc caggtcagtg tgcgggcctt ccacgctgcc agcggaacac tgga atg		177
		Met
gcg gaa ggg gaa cgg gtc tgc gcg tct gtk gtt ccc agc gct ctg cga		225
Ala Glu Gly      -60	-55	-50
acg ctg aaa agg agc aac ctg tcc aga atc ccc gca gga cag gaa		273
Thr Leu Lys Arg Arg Ser Asn Leu Ser Arg Ile Pro Ala Gly Gln Glu	-45	-40
		-35
aag gag ggg aaa tct cga cat gtt gct ccc cct ttt cgc ttt ttc cct		321
Lys Glu Gly Lys Ser Arg His Val Ala Pro Pro Phe Arg Phe Phe Pro	-30	-25
		-20
ttt tcc ggt ttt ttg ttt ttt ggt ttt ctt ttt ccc gtc ttt tct ttc		369
Phe Ser Gly Phe Leu Phe Phe Gly Phe Leu Phe Pro Val Phe Ser Phe	-15	-10
		-5
ccc tcc		1
Pro Ser		375

```
<210> 126
<211> 437
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 223..435
```

```
<221> sig_peptide
<222> 223..261
<223> Von Heijne matrix
      score 8.39999961853027
      seq MFCLAAILASASA/QR
```

```
<221> misc_feature
<222> 404
<223> n=a, q, c or t
```

```

<400> 126
tcaataccca tgtgaacagt ttcgtggagg gttttaagta ttttccactg gctggettthg      60
gggtataagta cctttccttc ttctgtcggtt aaccacgcgc aggggagaaa actatgcccc      120
cgtgaaagtc ccactctgtt ttccggtggg gaatactgga gcttaacctc ttggaggggg      180
ttgttccata ccaagggtcc ttccgtaggt atttctaatt gg atg ttc tgc ctg      234
                                     Met Phe Cys Leu
                                     -10
gca gca att tta gcc tca gca tct gcc caa cgg ttt cct tct gcc ttt      282
Ala Ala Ile Leu Ala Ser Ala Ser Ala Gln Arg Phe Pro Ser Ala Phe
          -5              1              5

```

```

tct cct tca cct tty yga tgg ctt yrg car tgt aas act gcc acc tcc      330
Ser Pro Ser Pro Phe Xaa Trp Leu Xaa Gln Cys Xaa Thr Ala Thr Ser
      10              15              20
ttg ggt ttt trc act gtg tgy art aac tcc ata att tcc ttg tgg tat      378
Leu Gly Phe Xaa Thr Val Cys Xaa Asn Ser Ile Ile Ser Leu Trp Tyr
      25              30              35
tta ayg ggr gtt ccc cca gag gtt ang gaa ctc cct ttc ttt cca tat      426
Leu Xaa Gly Val Pro Pro Glu Val Xaa Glu Leu Pro Phe Phe Pro Tyr
      40              45              50              55
tgc agc atg gg
Cys Ser Met

```

<210> 127  
 <211> 304  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 24..302

<221> sig\_peptide  
 <222> 24..74  
 <223> Von Heijne matrix  
       score 8.39999961853027  
       seq TLLLLLSEALALT/QT

```

<400> 127
ctcaggactc agaggctggg atc atg gta gat gga acc ctc ctt tta ctc ctc      53
                        Met Val Asp Gly Thr Leu Leu Leu Leu Leu
                        -15              -10
tcg gaa gcc ctg gcc ctt acc car acc tgg gcg ggc tcc cac tcc tkr      101
Ser Glu Ala Leu Ala Leu Thr Gln Thr Trp Ala Gly Ser His Ser Xaa
      -5              1              5
aag tat ttc cac act tcc gtg tcc cgg mcc ggc cgc ggg gag ccc cgc      149
Lys Tyr Phe His Thr Ser Val Ser Arg Xaa Gly Arg Gly Glu Pro Arg
      10              15              20              25
ttc atc tct gtg ggc tac gtg gac gac acc cgg tca gag tat tgg gac      197
Phe Ile Ser Val Gly Tyr Val Asp Asp Thr Arg Ser Glu Tyr Trp Asp
      30              35              40
cgg gag aca cgg agc gcc agg gac acc gca cag att ttc cga gtg aac      245
Arg Glu Thr Arg Ser Ala Arg Asp Thr Ala Gln Ile Phe Arg Val Asn
      45              50              55
ctg cgg acg ctg cgc ggc tac tac aat cag agc gag gcc ggg tct cam      293
Leu Arg Thr Leu Arg Gly Tyr Tyr Asn Gln Ser Glu Ala Gly Ser Xaa
      60              65              70
acc ctg cag tg
Thr Leu Gln
      75

```

<210> 128  
 <211> 244  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 19..243

<221> sig\_peptide  
 <222> 19..99  
 <223> Von Heijne matrix  
       score 8.39999961853027

seq LVLSLISLSIAWS/MV

&lt;221&gt; misc\_feature

&lt;222&gt; 112

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 128

gcgattaggt tttaaatgt atg aat ttc agg ggg cca caa acg ttc agt ctt 51  
 Met Asn Phe Arg Gly Pro Gln Thr Phe Ser Leu

-25

-20

tca cac agc ctt gtg tta tcc cta atc agt ctc tcc att gca tgg tct 99  
 Ser His Ser Leu Val Leu Ser Leu Ile Ser Leu Ser Ile Ala Trp Ser

-15

-10

-5

atg gtc gaa atg nbc act tct gca agc tac aar caa aag ttt gcc ctt 147  
 Met Val Glu Met Xaa Thr Ser Ala Ser Tyr Lys Gln Lys Phe Ala Leu

1

5

10

15

aga atc cta gtt gtg cag ttg ccc aca tgg gtg gaa tgt cca gta aac 195  
 Arg Ile Leu Val Val Gln Leu Pro Thr Trp Val Glu Cys Pro Val Asn

20

25

30

cac agg tgt gca cta ggg aga aag aat tgt tct att agg acc cag cca c 244  
 His Arg Cys Ala Leu Gly Arg Lys Asn Cys Ser Ile Arg Thr Gln Pro

35

40

45

&lt;210&gt; 129

&lt;211&gt; 232

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 156..230

&lt;221&gt; sig\_peptide

&lt;222&gt; 156..215

&lt;223&gt; Von Heijne matrix

score 8.39999961853027

seq SCICLFLPSLIHS/FP

&lt;400&gt; 129

ctacaggaag gaaaagtgtg acagctttga aaaagaaaga gggtaaaata tttaaccac 60  
 ccttggtgtc atttgtggca gcctatagca ttagagcctt tgagaacaga tctttccaga 120  
 ttctgcttaa gtccagggat tctgtgaccg cagaa atg act ggc atc tcc atc 173

Met Thr Gly Ile Ser Ile

-20

-15

tgc tcg tgc atc tgt ttg ttt ctt cct tca ttg att cac tca ttc ccc 221  
 Cys Ser Cys Ile Cys Leu Phe Leu Pro Ser Leu Ile His Ser Phe Pro

-10

-5

1

ccg ccc tgc gg 232  
 Pro Pro Cys

5

&lt;210&gt; 130

&lt;211&gt; 312

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 17..310

&lt;221&gt; sig\_peptide

&lt;222&gt; 17..94

<223> Von Heijne matrix  
 score 8.30000019073486  
 seq FLLLVAAAPRWVQL/QE

<400> 130  
 atgctttctg agagtc atg gac ctc ctg tgc aag aac atg aag cac ctg tgg 52  
                   Met Asp Leu Leu Cys Lys Asn Met Lys His Leu Trp  
                   -25                  -20                  -15  
 ttc ttc ctc ctg ctg gtg gcg gct ccc aga tgg gtc cag ctg cag gag 100  
 Phe Phe Leu Leu Leu Val Ala Ala Pro Arg Trp Val Gln Leu Gln Glu  
                   -10                  -5                  1  
 tcg ggc cca cgc ctg gtg agg cct ccg gag acc ctg aag cct tcg gag 148  
 Ser Gly Pro Arg Leu Val Arg Pro Pro Glu Thr Leu Lys Pro Ser Glu  
                   5                  10                  15  
 acc ctg tcc ctc acc tgc act att tct ggt gac tcc atg agc agt gct 196  
 Thr Leu Ser Leu Thr Cys Thr Ile Ser Gly Asp Ser Met Ser Ser Ala  
                   20                  25                  30  
 tct tac tat tgg gcc tgg atc cgc cag ccc cca ggc aag ggc ctg gaa 244  
 Ser Tyr Tyr Trp Ala Trp Ile Arg Gln Pro Pro Gly Lys Gly Leu Glu  
                   35                  40                  45                  50  
 ttc att ggg cgt gcc tta tat agt ggg acc acc gac tac aat ccg tcc 292  
 Phe Ile Gly Arg Ala Leu Tyr Ser Gly Thr Thr Asp Tyr Asn Pro Ser  
                   55                  60                  65  
 ctc agc agt cga atc acc ct 312  
 Leu Ser Ser Arg Ile Thr  
                   70

<210> 131  
 <211> 276  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 119..274

<221> sig\_peptide  
 <222> 119..253  
 <223> Von Heijne matrix  
 score 8.19999980926514  
 seq PLSLSLCLSLCHT/HT

<400> 131  
 gccttcacatct ctccatctct gcgctgctgc cggctgcgcc atccagcacc cagactccag 60  
 caccggccga ggacccccac tccggctgca gggaccctgt cccagcgaga ccgcaggc 118  
 atg tca tcc gaa aag tca gga ctc cca gac tca gtc cct cac act tct 166  
 Met Ser Ser Glu Lys Ser Gly Leu Pro Asp Ser Val Pro His Thr Ser  
                   -45                  -40                  -35                  -30  
 ccg ccg ccc tac aat gcc cct cag cct cca gcc gaa ccc cca gcc ccg 214  
 Pro Pro Pro Tyr Asn Ala Pro Gln Pro Pro Ala Glu Pro Pro Ala Pro  
                   -25                  -20                  -15  
 cct ctc tct ctc tct ctc tgt ctc tct ctc tgt cac aca cac aca cac 262  
 Pro Leu Ser Leu Ser Leu Cys Leu Ser Leu Cys His Thr His Thr His  
                   -10                  -5                  1  
 aca cac aca cac ac 276  
 Thr His Thr His  
                   5

<210> 132  
 <211> 174  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 35..172

<221> sig\_peptide  
 <222> 35..118  
 <223> Von Heijne matrix  
 score 8.19999980926514  
 seq LVSLLMQPEGALG/EE

<400> 132  
 actctgctga gctcctctgc acctgcccag gacc atg acg cct gct ctg cgc tgc 55  
 Met Thr Pro Ala Leu Arg Cys  
 -25  
 gca ttc gct ctg gcc ata gcg ggc ctc gtg tcg ctg ctg atg cag ccc 103  
 Ala Phe Ala Leu Ala Ile Ala Gly Leu Val Ser Leu Leu Met Gln Pro  
 -20 -15 -10  
 gag ggc gcc ctc ggc gag gag gct gca agt gcc gca gcc cag ggc cgc 151  
 Glu Gly Ala Leu Gly Glu Ala Ala Ser Ala Ala Ala Gln Gly Arg  
 -5 1 5 10  
 cag ttg gct gaa ctt agg ctc ca 174  
 Gln Leu Ala Glu Leu Arg Leu  
 15

<210> 133  
 <211> 344  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 133..342  
 <221> sig\_peptide  
 <222> 133..246  
 <223> Von Heijne matrix  
 score 8.19999980926514  
 seq LLLIFLSFPYTLC/IL

<400> 133  
 gcctttcact tgcacaaaca ctgttattat gatcacttat ccaactgaca tttttcagac 60  
 cttttaactt caactgttct tttttcctgt aaatcttaat tttctttttt tttctcccaa 120  
 tttttctcct ac atg tct gga ctc ttc cca gtt cct gtc aga gta aat gtt 171  
 Met Ser Gly Leu Phe Pro Val Pro Val Arg Val Asn Val  
 -35 -30  
 gat att gcc cag aac ata act tgc tct tcc ttt tct ctc ctt ctc att 219  
 Asp Ile Ala Gln Asn Ile Thr Cys Ser Ser Phe Ser Leu Leu Leu Ile  
 -25 -20 -15 -10  
 ttt ctt tct ttc ccc tac acc ctc tgt ata ctc tat aga gta aaa tca 267  
 Phe Leu Ser Phe Pro Tyr Thr Leu Cys Ile Leu Tyr Arg Val Lys Ser  
 -5 1 5  
 tat aca ccc acg gag tca ata act gcc ttt aat cta aca att ggg wga 315  
 Tyr Thr Pro Thr Glu Ser Ile Thr Ala Phe Asn Leu Thr Ile Gly Xaa  
 10 15 20  
 ttc cca tat ctt taw wtt tcw acc ccg gg 344  
 Phe Pro Tyr Leu Xaa Xaa Ser Thr Pro  
 25 30

<210> 134  
 <211> 244  
 <212> DNA  
 <213> Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 128..244

&lt;221&gt; sig\_peptide

&lt;222&gt; 128..226

&lt;223&gt; Von Heijne matrix

score 8.19999980926514

seq HALSLCLCTCAFA/FL

&lt;400&gt; 134

```

aagcaagaga ggggtgttca ggatgataaa gtcctggttg atgaaggcag atgcctgcag      60
ctcttccttg gggcagggct ggcttcata ggggtgcttg ttgggccctt tggaaggggg      120
tgtgcgg atg tgc agg gct gct tgt atc att aga atg gct gtt aga att      169
          Met Cys Arg Ala Ala Cys Ile Ile Arg Met Ala Val Arg Ile
                    -30                -25                -20
tca ttc ttt ctt tct tac cat gct ctg tct ctc tgc ctt tgt aca tgt      217
Ser Phe Phe Leu Ser Tyr His Ala Leu Ser Leu Cys Leu Cys Thr Cys
          -15                -10                -5
gcg ttt gca ttt ctc tcc ctc ctc ggg      244
Ala Phe Ala Phe Leu Ser Leu Leu Gly
          1                5

```

&lt;210&gt; 135

&lt;211&gt; 217

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 40..216

&lt;221&gt; sig\_peptide

&lt;222&gt; 40..90

&lt;223&gt; Von Heijne matrix

score 8.19999980926514

seq LLXALGFLXQVNP/XP

&lt;400&gt; 135

```

attaaaccac caccagstcc ccaagccacc ccttcagcc atg aag ttc ctg ctc      54
                                Met Lys Phe Leu Leu
                                           -15
ctg gma gcc ctc gga ttc ctg amc cag gtg aat ccc arc cca att sma      102
Leu Xaa Ala Leu Gly Phe Leu Xaa Gln Val Asn Pro Xaa Pro Ile Xaa
          -10                -5                1
ggd ggg tca aaa atg tgt gag twa cac ccc agg ata ctg cag gac atg      150
Gly Gly Ser Lys Met Cys Glu Xaa His Pro Arg Ile Leu Gln Asp Met
          5                10                15                20
ttg cca ctg ggg gga gac agc att gtt cat gtg caa cgc tks cag aaa      198
Leu Pro Leu Gly Gly Asp Ser Ile Val His Val Gln Arg Xaa Gln Lys
          25                30                35
atg ctg cat cag yta ctc c      217
Met Leu His Gln Leu Leu
          40

```

&lt;210&gt; 136

&lt;211&gt; 428

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 114..428

&lt;221&gt; sig\_peptide

&lt;222&gt; 114..239

&lt;223&gt; Von Heijne matrix

score 8.10000038146973

seq LFCFLLCLLSAAS/LL

&lt;400&gt; 136

```

aggcgtctgt gtgcgccgcc aagtcgggtgg ggcgggggacg cgaggtgtgg atgggggggtc   60
gccttgacct ctgcctcagc cagtagcgca gtctcggcct cgccgttacg gag atg   116
                                     Met
gtg ccc tgg gtg cgg acg atg ggg cag aag ctg aag cag cgg ctg cga   164
Val Pro Trp Val Arg Thr Met Gly Gln Lys Leu Lys Gln Arg Leu Arg
   -40                               -35                               -30
ctg gac gtg gga cgc gag atc tgc cgc cag tac ccg ctg ttc tgc ttc   212
Leu Asp Val Gly Arg Glu Ile Cys Arg Gln Tyr Pro Leu Phe Cys Phe
   -25                               -20                               -15                               -10
ctg ctg ctc tgt ctc agc gcc gcc tcc ctg ctt ctt aac agg tat att   260
Leu Leu Leu Cys Leu Ser Ala Ala Ser Leu Leu Leu Asn Arg Tyr Ile
                               -5                               1                               5
cat att tta atg atc ttc tgg tca ttt gtt gct gga gtt gtc aca ttt   308
His Ile Leu Met Ile Phe Trp Ser Phe Val Ala Gly Val Val Thr Phe
   10                               15                               20
tac tgc tca cta gga cct gat tct ctc tta cca aat ata ttc ttc aca   356
Tyr Cys Ser Leu Gly Pro Asp Ser Leu Leu Pro Asn Ile Phe Phe Thr
   25                               30                               35
ata aaa tac aaa ccc aag cag tta gga ctt cag gaa tta ttt cct caa   404
Ile Lys Tyr Lys Pro Lys Gln Leu Gly Leu Gln Glu Leu Phe Pro Gln
   40                               45                               50                               55
ggg cat agc tgt gct gtt tgt ggt   428
Gly His Ser Cys Ala Val Cys Gly
                               60

```

&lt;210&gt; 137

&lt;211&gt; 434

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 305..433

&lt;221&gt; sig\_peptide

&lt;222&gt; 305..406

&lt;223&gt; Von Heijne matrix

score 8.10000038146973

seq LLCLFXLFFFSFL/KR

&lt;400&gt; 137

```

tttctgtcca ttctccctca ccaccctgac gcaggtcttg ggaatgtgct gaaggtgcag   60
cagctgtctc acatttgtag cgaacacttt gactccaaag agaaggagga agacaaagac   120
aagaaggaaa agaaagacaa ggacaagaag gaagcccctg ctgacatggg agcacatcag   180
ggagtggctg ttctggggat tgcccttatt gctatggggg aggagattgg tgcagagatg   240
gcattacgaa cctttggcca cttggtgagt atagcatgaa gaaaattgga atatactggt   300
tttg atg gcc tgg ggt tcc cca ggg aag att ttt ctg atg ggt ttt ctt   349
      Met Ala Trp Gly Ser Pro Gly Lys Ile Phe Leu Met Gly Phe Leu
            -30                               -25                               -20
ggg gga gag ctg gtc ttt ttg ctg tgc ctt ttc ttw ctt ttt ttc ttt   397
Gly Gly Glu Leu Val Phe Leu Leu Cys Leu Phe Xaa Leu Phe Phe Phe
            -15                               -10                               -5
tct ttt ttg aag cgg agt ttt gct cta gag tgc aat g   434
Ser Phe Leu Lys Arg Ser Phe Ala Leu Glu Cys Asn
      1                               5

```

<210> 138  
 <211> 395  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 310..393

<221> sig\_peptide  
 <222> 310..357  
 <223> Von Heijne matrix  
 score 8.10000038146973  
 seq SILLLLAPPLPSA/VS

<221> misc\_feature  
 <222> 189  
 <223> n=a, g, c or t

<400> 138  
 aaaagctctg taaacatata ataaatggaa ttccattgac attcaagcct tacgtatttc 60  
 cagagcttct tcgacttata ctgcctcccc tactttaatt ctgttaaagt agttgaacac 120  
 cattcttctc ataatagttc tccctcsatt cttcagtgat tyccttgtgt ttataggata 180  
 aagtccacnt gttatttttg cagtcagttc aagatccaca aatcagtcct tacccttaca 240  
 tccttatttc tcactgctgt tctaataatag tctttatacc agtcaggctg gtctgttcac 300  
 tattcctga atg ttt ttc tcc att ctt ttg tta ttg gca ccc ccg cta ccc 351  
 Met Phe Phe Ser Ile Leu Leu Leu Leu Ala Pro Pro Leu Pro  
 -15 -10 -5  
 tct gca gtg tct ttg cta cct ttc ttt ttc tac tgt gtg cag gg 395  
 Ser Ala Val Ser Leu Leu Pro Phe Phe Phe Tyr Cys Val Gln  
 1 5 10

<210> 139  
 <211> 268  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 141..266

<221> sig\_peptide  
 <222> 141..206  
 <223> Von Heijne matrix  
 score 8.10000038146973  
 seq LLVCSWLSISLHA/HT

<400> 139  
 caactctgct gttttgtagg aagccacatg gaggtcattt acggttacta gttatcttag 60  
 tcagcttggg cagccattaa aaaataatac tgtagacgga gtggcccaaa cgagagaaat 120  
 ttatttctta tagttttggc atg gta gat ttc atc ctg agg tct ctt ctc ttg 173  
 Met Val Asp Phe Ile Leu Arg Ser Leu Leu Leu  
 -20 -15  
 gtt tgt agt tgg ctg tca atc tcc ctg cat gct cac acg acc gct ttt 221  
 Val Cys Ser Trp Leu Ser Ile Ser Leu His Ala His Thr Thr Ala Phe  
 -10 -5 1 5  
 tgt aca tac agt aag aaa ata cac act gtc atg tca ttt ttt tgt aa 268  
 Cys Thr Tyr Ser Lys Lys Ile His Thr Val Met Ser Phe Phe Cys  
 10 15 20

<210> 140



<211> 170  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 93..170

<221> sig\_peptide  
 <222> 93..140  
 <223> Von Heijne matrix  
 score 8.10000038146973  
 seq LLYFLCVSSYVTS/FF

<400> 140  
 ttttgactga tatcaaattc taggtggacc gagattttct ttcagtcttt caaagatatt 60  
 actctattgc cttctatctt gcatagtttc tg atg aga agt ctg ttg tat ttc 113  
 Met Arg Ser Leu Leu Tyr Phe  
 -15 -10  
 tta tgt gtt tct tca tat gta aca tct ttt ttc ttt ttt ttt ttt 161  
 Leu Cys Val Ser Ser Tyr Val Thr Ser Phe Phe Phe Phe Phe Phe  
 -5 1 5  
 ttt ttt ttt 170  
 Phe Phe Phe  
 10

<210> 141  
 <211> 396  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 192..395

<221> sig\_peptide  
 <222> 192..236  
 <223> Von Heijne matrix  
 score 8  
 seq FISFLCLIALAGT/SS

<400> 141  
 gattctcagc ttagttgctg ttggtgtata ggagagctac tgatttgtgt acattaattt 60  
 tgtatccgga aactttgttg aattatttta tcagtcttag gagctttttg gaggagtctt 120  
 tagggttctc taggtataca atcatatcat cagcaaacag tgacaattcg acttcctctt 180  
 tatggatttg t atg ccc ttt att tct ttc ctt tgt ctg att gct ctg gct 230  
 Met Pro Phe Ile Ser Phe Leu Cys Leu Ile Ala Leu Ala  
 -15 -10 -5  
 ggg act tcc agt act atg ttg aga agt gct ctg gct ggg act tcc agt 278  
 Gly Thr Ser Ser Thr Met Leu Arg Ser Ala Leu Ala Gly Thr Ser Ser  
 1 5 10  
 act atg tkg arg aga agt ggt gam agt ggg wat cct kgh ctk gty cma 326  
 Thr Met Xaa Xaa Arg Ser Gly Xaa Ser Gly Xaa Pro Xaa Leu Val Xaa  
 15 20 25 30  
 gtc ctm aga ggg aat gct ttc agc ttt ttc cca ttc agt ctg atg twg 374  
 Val Leu Arg Gly Asn Ala Phe Ser Phe Phe Pro Phe Ser Leu Met Xaa  
 35 40 45  
 gct atg ggt tgt cat aga tgg c 396  
 Ala Met Gly Cys His Arg Trp  
 50

<210> 142  
 <211> 357

<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 292..357

<221> sig\_peptide  
<222> 292..339  
<223> Von Heijne matrix  
score 8  
seq FLLGAIFIALSSS/RI

<400> 142  
cgtgcctgcg caatgggtgt cgggtccgct ttttcccaat ccggacgtaa tcgtgggtttt 60  
tgttctgcaa taggcggctt agagggaggg gctttttcgc ctataacctac tgtagcttct 120  
ccacgtatgg accctaaagg ctactgctgc tactacgggg ctagacagtt actgtctcag 180  
ctctaggatg tgcgttcttc cactagaagc tcttctgagg gaggtaatta aaaaacagtg 240  
gaatggaaaa acagtgtctgt agtcacctcg taatatgctc cttgtcaaca a atg tat 297  
Met Tyr  
-15  
aca ttc ctg cta ggt gcc ata ttc att gct tta agc tca agt cgc atc 345  
Thr Phe Leu Leu Gly Ala Ile Phe Ile Ala Leu Ser Ser Ser Arg Ile  
-10 -5 1  
tta cta gtg aag 357  
Leu Leu Val Lys  
5

<210> 143  
<211> 159  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 26..157

<221> sig\_peptide  
<222> 26..151  
<223> Von Heijne matrix  
score 7.90000009536743  
seq LVCVCVCVCVCXC/XR

<400> 143  
tgtgtgtgtg tgtgtctgcg tgtgt atg tgt ttg tgt ccc tgc tgg gat gtg 52  
Met Cys Leu Cys Pro Cys Trp Asp Val  
-40 -35  
ttt act gtg ttt gtg tgt gtc tct gtg tgt gtg tct gtg tct gtc cct 100  
Phe Thr Val Phe Val Cys Val Ser Val Cys Val Ser Val Ser Val Pro  
-30 -25 -20  
gtc ggg atg tat tta gtg tgt gtg tgt gtg tgt gtg tgt stc 148  
Val Gly Met Tyr Leu Val Cys Val Cys Val Cys Val Cys Val Cys Xaa  
-15 -10 -5  
tgc gyg cgt gg 159  
Cys Xaa Arg  
1

<210> 144  
<211> 433  
<212> DNA  
<213> Homo sapiens

<220>

<221> CDS  
<222> 282..431

<221> sig\_peptide  
<222> 282..383  
<223> Von Heijne matrix  
score 7.90000009536743  
seq LFSLLMLTQSPLA/GQ

<221> misc\_feature  
<222> 132,149  
<223> n=a, g, c or t

<400> 144  
aaaataagggt atctggcaaa agaatatatg aaagagtatg aagaactctc cttgaaagct 60  
gtggccccc ttggccatgg ctgcagagcc gatgtcccg ccaatccagg cgggatcccc 120  
ttgaagcmgg knsmwhbcty kragscwknc cmabtctccg ggggcaastc tttcccttc 180  
cctgtgaccc kcttcggaca gttgaccatc tcaacaccta gtgggttaaaa agaagagcat 240  
ggacggcctg gggcctgcac tggctgtgct gggagtttgt c atg ttg ata gct aag 296  
Met Leu Ile Ala Lys  
-30  
cag gcc cag ccc caa ggc ctc act gcc atc tgc ttc cct ctc aca cct 344  
Gln Ala Gln Pro Gln Gly Leu Thr Ala Ile Cys Phe Pro Leu Thr Pro  
-25 -20 -15  
ctc ttc tcc ctc ctc atg ctc act cag agc ccc ctt gca ggt cag gaa 392  
Leu Phe Ser Leu Leu Met Leu Thr Gln Ser Pro Leu Ala Gly Gln Glu  
-10 -5 1  
gga aga gaa gga ggg aaa gaa cgg tac ttg ttg gtg att ca 433  
Gly Arg Glu Gly Gly Lys Glu Arg Tyr Leu Leu Val Ile  
5 10 15

<210> 145  
<211> 200  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 15..200

<221> sig\_peptide  
<222> 15..92  
<223> Von Heijne matrix  
score 7.90000009536743  
seq RVCLLSLSLFLWA/NR

<400> 145  
aatacgccag gaac atg cta agg acc tgg agc tct cta ccc tgg acc cgt 50  
Met Leu Arg Thr Trp Ser Ser Leu Pro Trp Thr Arg  
-25 -20 -15  
ttt cgg gtt tgc ttg ctc tct ctc tct ctc ttt ctc tgg gct aat cgt 98  
Phe Arg Val Cys Leu Leu Ser Leu Ser Leu Phe Leu Trp Ala Asn Arg  
-10 -5 1  
tta gag gac agt cgc tcc tgc caa cct aat ccc atg agc ctg act acc 146  
Leu Glu Asp Ser Arg Ser Cys Gln Pro Asn Pro Met Ser Leu Thr Thr  
5 10 15  
ttg ccg ggc cac agg ctc aaa gaa gca gtg tgg ctg cca gca ccc tca 194  
Leu Pro Gly His Arg Leu Lys Glu Ala Val Trp Leu Pro Ala Pro Ser  
20 25 30  
ctt ggg 200  
Leu Gly  
35

<210> 146  
 <211> 297  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 80..295

<221> sig\_peptide  
 <222> 80..166  
 <223> Von Heijne matrix  
 score 7.90000009536743  
 seq LVVXWLLPXQCSC/ER

<400> 146  
 aacacccag cccaagtcca tccccggtcc cttggcagca gtgcgcatcc acaaagccag 60  
 cggcacaatt taattactg atg gcc cct ttc cta cga cag gtg gat rtg tgg 112  
 Met Ala Pro Phe Leu Arg Gln Val Asp Xaa Trp  
 -25 -20  
 gga gca cag gcc ggt ctg gtg gtb gsm tgg tta cta cca tgs caa tgc 160  
 Gly Ala Gln Ala Gly Leu Val Val Xaa Trp Leu Leu Pro Xaa Gln Cys  
 -15 -10 -5  
 agc tgt gaa cga tca gag caa tat ctg agc acc tgt ctc cca cag cac 208  
 Ser Cys Glu Arg Ser Glu Gln Tyr Leu Ser Thr Cys Leu Pro Gln His  
 1 5 10  
 tca agc atc aag cag tgc tgc atc aag cat cca gca ggc ccg atc ccc 256  
 Ser Ser Ile Lys Gln Ser Cys Ile Lys His Pro Ala Gly Pro Ile Pro  
 15 20 25 30  
 gca ggc cac cta cag gga aag gcc aca gct gcg ccc ctg gg 297  
 Ala Gly His Leu Gln Gly Lys Ala Thr Ala Ala Pro Leu  
 35 40

<210> 147  
 <211> 300  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 80..298

<221> sig\_peptide  
 <222> 80..136  
 <223> Von Heijne matrix  
 score 7.90000009536743  
 seq WLFLVAILKGVRC/EV

<400> 147  
 agctctgaga gaggagccca gccctgggat cttcaggtgt tttcactcgg tgatcaggac 60  
 tgcacagaga gaactcacc atg gag ttt ggg ctg aag tgg ctt ttt ctt gtg 112  
 Met Glu Phe Gly Leu Lys Trp Leu Phe Leu Val  
 -15 -10  
 gca att tta aaa ggt gtc cgg tgt gaa gtg aag ctg gtg gag tct ggg 160  
 Ala Ile Leu Lys Gly Val Arg Cys Glu Val Lys Leu Val Glu Ser Gly  
 -5 1 5  
 gga ggc ctg gtg cag ccg ggg ggg tcc ctg aga ctc tcc tgt gta gga 208  
 Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys Val Gly  
 10 15 20  
 tct gga ttc gtc ttc gat aaa tat ggc ata agt tgg gtg cgc cag gca 256  
 Ser Gly Phe Val Phe Asp Lys Tyr Gly Ile Ser Trp Val Arg Gln Ala  
 25 30 35 40

cca gga aag ggc cta cag tgg gtc gcg ggg atc ggt ggc ggg gg 300  
 Pro Gly Lys Gly Leu Gln Trp Val Ala Gly Ile Gly Gly Gly  
                     45                    50

<210> 148  
 <211> 405  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 21..404

<221> sig\_peptide  
 <222> 21..68  
 <223> Von Heijne matrix  
       score 7.90000009536743  
       seq AMLVLVVSPWSAA/RG

<400> 148  
 gcggtcttcc agcagggaaa atg gcg ctg gcc atg ctg gtc ttg gtg gtt tcg 53  
                     Met Ala Leu Ala Met Leu Val Leu Val Val Ser  
                     -15                    -10  
 ccg tgg tct gcg gcc cgg gga gtg ctt cga aac tac tgg gag cga ctg 101  
 Pro Trp Ser Ala Ala Arg Gly Val Leu Arg Asn Tyr Trp Glu Arg Leu  
 -5                    1                    5                    10  
 cta cgg aag ctt ccg cag agc cgg ccg ggc ttt ccc agt cct ccg tgg 149  
 Leu Arg Lys Leu Pro Gln Ser Arg Pro Gly Phe Pro Ser Pro Pro Trp  
                     15                    20                    25  
 gga cca gca tta gca gta cag ggc cca gcc atg ttt aca gag cca gca 197  
 Gly Pro Ala Leu Ala Val Gln Gly Pro Ala Met Phe Thr Glu Pro Ala  
                     30                    35                    40  
 aat gat acc agt gga agt aaa gag aat tcc agc ctt ttg gac agt atc 245  
 Asn Asp Thr Ser Gly Ser Lys Glu Asn Ser Ser Leu Leu Asp Ser Ile  
                     45                    50                    55  
 ttt tgg atg gca gct ccc aaa aat aga cgc acc att gaa gtt aac cgg 293  
 Phe Trp Met Ala Ala Pro Lys Asn Arg Arg Thr Ile Glu Val Asn Arg  
 60                    65                    70                    75  
 tgt agg aga aga aat ccg cag aag ctt att aaa gtt aag aac aac ata 341  
 Cys Arg Arg Arg Asn Pro Gln Lys Leu Ile Lys Val Lys Asn Asn Ile  
                     80                    85                    90  
 gac gtt tgt cct gaa tgt ggt cac ctg aaa cag aaa srt gtc ctt tgt 389  
 Asp Val Cys Pro Glu Cys Gly His Leu Lys Gln Lys Xaa Val Leu Cys  
                     95                    100                    105  
 gct act gct atg aaa a 405  
 Ala Thr Ala Met Lys  
                     110

<210> 149  
 <211> 146  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 56..145

<221> sig\_peptide  
 <222> 56..115  
 <223> Von Heijne matrix  
       score 7.80000019073486  
       seq LLLFPLSLLFTLG/FL

&lt;400&gt; 149

```

aaaccttctg actactaacc tgtagatccc tttagttcct tagcagtatt caciaa atg      58
                                     Met
                                     -20
ttt ttc tac tca cac ttt tta ctt ctt ttt ccc ctc tcg tta ctt ttc      106
Phe Phe Tyr Ser His Phe Leu Leu Leu Phe Pro Leu Ser Leu Leu Phe
          -15                      -10                      -5
aca ctt gga ttt ttg ttt gtc ttt ttt ttt ttt ttt t      146
Thr Leu Gly Phe Leu Phe Val Phe Phe Phe Phe Phe Phe
          1                      5                      10

```

&lt;210&gt; 150

&lt;211&gt; 408

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 105..407

&lt;221&gt; sig\_peptide

&lt;222&gt; 105..242

&lt;223&gt; Von Heijne matrix

score 7.80000019073486

seq LVLLGTRVPLSGG/GP

&lt;400&gt; 150

```

aaacagggcc attggcaaag ctgggggtacc agtcacccag ccacgctcta ggggtggtagc      60
caagaagacg gaccccgagt gggaggcaga gagacaagag gtgg atg aag cag agc      116
                                     Met Lys Gln Ser
                                     -45
aag cgt gas atg gtg aag aga aga cgg agc ccc gcg ctg gga gag gaa      164
Lys Arg Xaa Met Val Lys Arg Arg Arg Ser Pro Ala Leu Gly Glu Glu
          -40                      -35                      -30
cgc ttc agt ccg agt tcc att ctg cac cca agg ctc ccc ttg gtc ctc      212
Arg Phe Ser Pro Ser Ser Ile Leu His Pro Arg Leu Pro Leu Val Leu
          -25                      -20                      -15
ctg gga acc agg gtg ccc ctt agt ggt ggt ggc cca gga gaa ccc gac      260
Leu Gly Thr Arg Val Pro Leu Ser Gly Gly Gly Pro Gly Glu Pro Asp
          -10                      -5                      1                      5
caa ggc agg agc gcc ccc tcc tgg aag agc ctc gct tca acg cat mat      308
Gln Gly Arg Ser Ala Pro Ser Trp Lys Ser Leu Ala Ser Thr His Xaa
          10                      15                      20
cat tcc cgg ccg gca gca ggg gcg acg cca gca agg cct gcg act cag      356
His Ser Arg Pro Ala Ala Gly Ala Thr Pro Ala Arg Pro Ala Thr Gln
          25                      30                      35
agc cag ctt ggc ccg ttc gcc ccg ccc ctt ccc ggt gtc cgc ccc gcc      404
Ser Gln Leu Gly Pro Phe Ala Pro Pro Leu Pro Gly Val Arg Pro Ala
          40                      45                      50
cca t
Pro
55

```

&lt;210&gt; 151

&lt;211&gt; 166

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 69..164

&lt;221&gt; sig\_peptide

&lt;222&gt; 69..122

&lt;223&gt; Von Heijne matrix

score 7.80000019073486

seq LCVLSLLVSFKSA/CL

&lt;400&gt; 151

cacattttct acttaaaagc aamgttacaa agcctgtgga attgctctga cttagaaaga 60

acttgatc atg ctt ttg gag tct cta tgt gtt ctc tct ctg ttg gtt agt 110

Met Leu Leu Glu Ser Leu Cys Val Leu Ser Leu Leu Val Ser

-15

-10

-5

ttt aaa tca gcc tgc ctc aca agg gag cct gca ttt gat tcc caa gcc 158

Phe Lys Ser Ala Cys Leu Thr Arg Glu Pro Ala Phe Asp Ser Gln Ala

1

5

10

cgc ccg gg

166

Arg Pro

&lt;210&gt; 152

&lt;211&gt; 382

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 99..380

&lt;221&gt; sig\_peptide

&lt;222&gt; 99..236

&lt;223&gt; Von Heijne matrix

score 7.80000019073486

seq LLYLSFAALGVVA/LR

&lt;400&gt; 152

ttttacacac acacatacat acacacacac agctaattga gttttaaagt aatattcttg 60

ctaatacccta ctgaattgta gcttggtggt gtttctga atg gtt ttt gga tat tgg 116

Met Val Phe Gly Tyr Trp

-45

aag cag ccg ctg att acc ctt gca aag aaa tct gta aaa tgt gca cgt 164

Lys Gln Pro Leu Ile Thr Leu Ala Lys Lys Ser Val Lys Cys Ala Arg

-40

-35

-30

-25

gaa tgt ctg aga tgc tct ctc agg cct cta gtc ctt ctg tat ctt tcc 212

Glu Cys Leu Arg Cys Ser Leu Arg Pro Leu Val Leu Leu Tyr Leu Ser

-20

-15

-10

ttt gca gcc ctg ggt gta gta gca ctc agg agt gtt gaa tca ccc ctg 260

Phe Ala Ala Leu Gly Val Val Ala Leu Arg Ser Val Glu Ser Pro Leu

-5

1

5

gcc gag acc cac tcc tgc tgg ctc agc ctg ggc atg tgt gtg ctc cag 308

Ala Glu Thr His Ser Cys Trp Leu Ser Leu Gly Met Cys Val Leu Gln

10

15

20

tgt gaa cag cag tgg gtt cca acc cca gtc tcc ttt ctc tgt ggc ctc 356

Cys Glu Gln Gln Trp Val Pro Thr Pro Val Ser Phe Leu Cys Gly Leu

25

30

35

40

tct ggc tcc agc acc atc atc gtt ag 382

Ser Gly Ser Ser Thr Ile Ile Val

45

&lt;210&gt; 153

&lt;211&gt; 208

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 10..207

&lt;221&gt; sig\_peptide

&lt;222&gt; 10..81

&lt;223&gt; Von Heijne matrix

score 7.80000019073486

seq CVYVVCLVSCVLC/VV

&lt;400&gt; 153

tgcaatgtc	atg	tgt	gtt	gtg	tgc	agt	gtg	cat	ggg	gtg	tgt	tgt	gta	tat	51
	Met	Cys	Val	Val	Cys	Ser	Val	His	Gly	Val	Cys	Cys	Val	Tyr	

-20

-15

gtg	gtg	tgc	ctg	gtg	tgc	tgt	gtt	ttg	tgt	gtc	gtg	tgt	cct	gtg	tgt	99
Val	Val	Cys	Leu	Val	Ser	Cys	Val	Leu	Cys	Val	Val	Cys	Pro	Val	Cys	

-10

-5

1

5

tgg	gtt	atg	tgt	tgt	gtg	tgg	tgc	atc	tgt	gtg	tgt	gtg	tgg	tgt	gtc	147
Trp	Val	Met	Cys	Cys	Val	Trp	Cys	Ile	Cys	Val	Cys	Val	Trp	Cys	Val	

10

15

20

tgt	tgt	atg	tgt	tgt	gtg	ttg	tca	tgt	gtt	gtg	tca	cat	ggg	ttg	tgt	195
Cys	Cys	Met	Cys	Cys	Val	Leu	Ser	Cys	Val	Val	Ser	His	Gly	Leu	Cys	

25

30

35

ggg	gtg	tca	tgg	g												208
Gly	Val	Ser	Trp													

40

&lt;210&gt; 154

&lt;211&gt; 251

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 73..249

&lt;221&gt; sig\_peptide

&lt;222&gt; 73..129

&lt;223&gt; Von Heijne matrix

score 7.80000019073486

seq WVFLVAVLEVVCQ/EI

&lt;400&gt; 154

agagaggagc	ctcagcccta	gactccaagg	cctttccact	tggtgatcag	cactgagcac	60									
agaggactca	cc	atg	gaa	ctg	ggg	ctg	tcc	tgg	gtc	ttc	ctt	gtt	gct	gtt	111

Met Glu Leu Gly Leu Ser Trp Val Phe Leu Val Ala Val

-15

-10

tta	gaa	gtt	gtc	cag	tgt	gaa	att	caa	ctg	att	gac	gcc	ggg	gga	ggc	159
Leu	Glu	Val	Val	Gln	Cys	Glu	Ile	Gln	Leu	Ile	Asp	Ala	Gly	Gly	Gly	

-5

1

5

10

cac	gtc	cag	gag	ggg	ggg	tca	ctg	aga	ctc	tcc	tgt	gtt	gcc	tct	gac	207
His	Val	Gln	Ala	Gly	Gly	Ser	Leu	Arg	Leu	Ser	Cys	Val	Ala	Ser	Asp	

15

20

25

ttc	ctg	ttt	aga	agc	tat	tgg	atg	acc	tgg	gtc	cgc	cat	ccg	gg		251
Phe	Leu	Phe	Arg	Ser	Tyr	Trp	Met	Thr	Trp	Val	Arg	His	Pro			

30

35

40

&lt;210&gt; 155

&lt;211&gt; 147

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 24..146





ttt att aag cct ctg gg  
Phe Ile Lys Pro Leu  
15

115

<210> 158  
<211> 175  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 54..173

<221> sig\_peptide  
<222> 54..131  
<223> Von Heijne matrix  
score 7.69999980926514  
seq FLLLLYFFXIAVT/HP

<400> 158  
caattcaaca tgagatttag tgggtgacaaa tatccaaact ctatcaacct cta atg 56  
Met  
ctg acc tca ctg cct ttc ctc ctg ccc acc atc agc ttt ctc ctc ctc 104  
Leu Thr Ser Leu Pro Phe Leu Leu Pro Thr Ile Ser Phe Leu Leu Leu  
-25 -20 -15 -10  
ttg tat ttt ttt cma att gct gtc acc cat ccg tca gtt ctc atc aac 152  
Leu Tyr Phe Phe Xaa Ile Ala Val Thr His Pro Ser Val Leu Ile Asn  
-5 1 5  
ttc tct ttc tcc ttc ccc aga tc 175  
Phe Ser Phe Ser Phe Pro Arg  
10

<210> 159  
<211> 230  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 121..228

<221> sig\_peptide  
<222> 121..180  
<223> Von Heijne matrix  
score 7.59999990463257  
seq LLFFTCGLPALHG/DS

<221> misc\_feature  
<222> 18  
<223> n=a, g, c or t

<400> 159  
aggaggggcc gtcagggngg gatacagcct ggaaggtgcg tgtggggctg ggtctcggag 60  
tgggagacgt ggagtgacag taatgcatgt ccatgggtaca caaattcaca aggtttgtaa 120  
atg aga aaa gac gtg agg ttc ctt ttg ttc ttt acc tgt ggc ctc cct 168  
Met Arg Lys Asp Val Arg Phe Leu Leu Phe Phe Thr Cys Gly Leu Pro  
-20 -15 -10 -5  
gcc cta cac ggg gac tct agg gtg gaa tgt agc aaa gcc cat cca cca 216  
Ala Leu His Gly Asp Ser Arg Val Glu Cys Ser Lys Ala His Pro Pro  
1 5 10  
gcc atg tac tac cc 230  
Ala Met Tyr Tyr

15

<210> 160  
 <211> 346  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 202..345

<221> sig\_peptide  
 <222> 202..282  
 <223> Von Heijne matrix  
 score 7.59999990463257  
 seq WTLLSISLSVFWS/EP

<400> 160  
 ttcttctaca tacagctacc caactagccc acaccattta ttgaatacag agtagtcttt 60  
 tcctgtgttg ttatttttct taactttggt aaagatcaga tatctgtagg tgtgcagctt 120  
 tatttctggg ttttctgttc cgttccattg gtctatgtgt ctgtttttgt accagtacca 180  
 tgctgttctg gcaccagtac c atg cta ttt tgg tta cca tct cca tct gag 231  
 Met Leu Phe Trp Leu Pro Ser Pro Ser Glu  
 -25 -20  
 acc act tca gcc tgg act tta ttg tcc ata tca cta tca gta ttt tgg 279  
 Thr Thr Ser Ala Trp Thr Leu Leu Ser Ile Ser Leu Ser Val Phe Trp  
 -15 -10 -5  
 tca gag cca ttc aat aag tct cta gga agt tcc aaa cta cca tgt cat 327  
 Ser Glu Pro Phe Asn Lys Ser Leu Gly Ser Ser Lys Leu Pro Cys His  
 1 5 10 15  
 ttt ttt tct ata aaa cgg g 346  
 Phe Phe Ser Ile Lys Arg  
 20

<210> 161  
 <211> 388  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 194..388

<221> sig\_peptide  
 <222> 194..334  
 <223> Von Heijne matrix  
 score 7.59999990463257  
 seq LXLGEGLTFLCLC/QV

<221> misc\_feature  
 <222> 352  
 <223> n=a, g, c or t

<400> 161  
 agtgagagct tagtcttggt actatttggt tttgtttctt actgtttgtc tgtttatggt 60  
 tgggtgcaag aaaattgtgt tgtaaattat ccttgcttt ctctattagt taatagcctt 120  
 ccccttctgt agtaaagtaa msagsctttt kcctgttcaa atattttagg cttgtttttt 180  
 gttttgattg tac atg cct gtg tgt ttt tat tcc tta att tgt ttc ttt 229  
 Met Pro Val Cys Phe Tyr Ser Leu Ile Cys Phe Phe  
 -45 -40  
 att tat ttc tgt ttg tta tct cca aga gaa aca ata gaa gag gtg gcc 277  
 Ile Tyr Phe Cys Leu Leu Ser Pro Arg Glu Thr Ile Glu Glu Val Ala

98

```

-35          -30          -25          -20
ctc ttc cag ttt tct ctg cth mtc ttg gga gag ggt ctc acc ttt ctt      325
Leu Phe Gln Phe Ser Leu Leu Xaa Leu Gly Glu Gly Leu Thr Phe Leu
          -15          -10          -5
tgc ctc tgc cag gta atg acg aat aan atg caa ctg ctg ttc ttg agt      373
Cys Leu Cys Gln Val Met Thr Asn Xaa Met Gln Leu Leu Phe Leu Ser
          1          5          10
ggg gta gtc tgt ggg      388
Gly Val Val Cys Gly
          15

```

<210> 162  
 <211> 235  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 172..234  
  
 <221> sig\_peptide  
 <222> 172..210  
 <223> Von Heijne matrix  
       score 7.5  
       seq MAPLLLSLSCSFS/CH

```

<400> 162
ccccccaaaa tctcatgttg agatttgatc cctaattgttg gagatggggc ctggtgggag      60
atattcggat catgagggca gatccctcac taatggcctg gtgccctccc tgtggaaatg      120
agtaagtctt cactcttttg gttcacctga gagctgtttg tttaaaagag c atg gca      177
                                   Met Ala
ccc ctc ctt ctc tct ctg tct tgc tcc ttt tct tgc cat gtg aca ctc      225
Pro Leu Leu Leu Ser Leu Ser Cys Ser Phe Ser Cys His Val Thr Leu
          -10          -5          1          5
ctg ccc cgg g      235
Leu Pro Arg

```

<210> 163  
 <211> 240  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 99..239  
  
 <221> sig\_peptide  
 <222> 99..158  
 <223> Von Heijne matrix  
       score 7.5  
       seq LLWVLLNLGPRA/AG

```

<400> 163
aaaacgaccc ggtgggtcta cagcgggaagg gagggagcga aggtaggagg cagggcttgc      60
ctcactggcc accctcccaa cccaagagc ccagcccc atg gtc ccc gcc gcc ggs      116
                                   Met Val Pro Ala Ala Gly
                                   -20          -15
gcg ctg ctg tgg gtc ctg ctg ctg aat ctg ggt ccc cgg gcg gcg ggg      164
Ala Leu Leu Trp Val Leu Leu Leu Asn Leu Gly Pro Arg Ala Ala Gly
          -10          -5          1
gcc caa ggc ctg acc cag act ccg acc gaa atg cag cgg gtc agt tta      212
Ala Gln Gly Leu Thr Gln Thr Pro Thr Glu Met Gln Arg Val Ser Leu
          5          10          15

```

99

cgc ttt ggg ggc ccc atg acc cgc agg g 240  
 Arg Phe Gly Gly Pro Met Thr Arg Arg  
 20 25

<210> 164  
 <211> 195  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 20..193

<221> sig\_peptide  
 <222> 20..91  
 <223> Von Heijne matrix  
 score 7.5  
 seq LISELLLLRSVTS/HN

<400> 164  
 ttctgattat gatggtaat atg gta ttc tgg gaa ata tct gtc caa att atc 52  
 Met Val Phe Trp Glu Ile Ser Val Gln Ile Ile  
 -20 -15  
 ctg atc tct gaa ctc ctg ctg ttg agg tca gtc act tca cac aat acc 100  
 Leu Ile Ser Glu Leu Leu Leu Leu Arg Ser Val Thr Ser His Asn Thr  
 -10 -5 1  
 atg atg aga gct tta tca agc cag atg ctt agt cag agc ttt cca aga 148  
 Met Met Arg Ala Leu Ser Ser Gln Met Leu Ser Gln Ser Phe Pro Arg  
 5 10 15  
 ccc agc ttt ggt ttt atc agc aaa atc cat cct tcc cac ccc ccc aa 195  
 Pro Ser Phe Gly Phe Ile Ser Lys Ile His Pro Ser His Pro Pro  
 20 25 30

<210> 165  
 <211> 256  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 34..255

<221> sig\_peptide  
 <222> 34..186  
 <223> Von Heijne matrix  
 score 7.5  
 seq VVSLTFLLGMTWG/FA

<221> misc\_feature  
 <222> 18  
 <223> n=a, g, c or t

<400> 165  
 tatttatgtg acctgtgngg gtattttgga gtc atg ttt ttt ctg aac att gcc 54  
 Met Phe Phe Leu Asn Ile Ala  
 -50 -45  
 atg ttc att gtg gta atg gtg cag atc tgt ggg agg aat ggc aag aga 102  
 Met Phe Ile Val Val Met Val Gln Ile Cys Gly Arg Asn Gly Lys Arg  
 -40 -35 -30  
 agc aac cgg acc ctg aga gaa gaa gtg tta agg aac ctg cgc agt gtg 150  
 Ser Asn Arg Thr Leu Arg Glu Glu Val Leu Arg Asn Leu Arg Ser Val  
 -25 -20 -15

100

```

gtt agc ttg acc ttt ctg ttg ggc atg aca tgg ggt ttt gca ttc ttt      198
Val Ser Leu Thr Phe Leu Leu Gly Met Thr Trp Gly Phe Ala Phe Phe
      -10                      -5                      1
gcc tgg gga ccc tta aat atc ccc ttc atg tac ctc ttc tcc atc ttc      246
Ala Trp Gly Pro Leu Asn Ile Pro Phe Met Tyr Leu Phe Ser Ile Phe
      5                      10                      15                      20
aat tca tta c
Asn Ser Leu

```

&lt;210&gt; 166

&lt;211&gt; 209

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 36..209

&lt;221&gt; sig\_peptide

&lt;222&gt; 36..86

&lt;223&gt; Von Heijne matrix

score 7.5

seq FLFLFLLXXLIVA/VT

&lt;400&gt; 166

```

cttttttdtc ckgcacaagg gatttccggg tcagg atg aac aaa cac ttc ttg      53
                                Met Asn Lys His Phe Leu
                                -15
ttc ctc ttc ctc ctt dac kgc ctc att gtg gca gtg aca tca ctt cag      101
Phe Leu Phe Leu Leu Xaa Xaa Leu Ile Val Ala Val Thr Ser Leu Gln
      -10                      -5                      1                      5
tgc ata aca tgc cac ctt cgc aca cgg aca gac cgc tgt aga aga ggc      149
Cys Ile Thr Cys His Leu Arg Thr Arg Thr Asp Arg Cys Arg Arg Gly
      10                      15                      20
ttt ggt gdc tgt act gct cag aag ggc gag gca tgc atg ctc tta agg      197
Phe Gly Xaa Cys Thr Ala Gln Lys Gly Glu Ala Cys Met Leu Leu Arg
      25                      30                      35
att cac cag cgc
Ile His Gln Arg
      40

```

&lt;210&gt; 167

&lt;211&gt; 184

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 44..184

&lt;221&gt; sig\_peptide

&lt;222&gt; 44..148

&lt;223&gt; Von Heijne matrix

score 7.5

seq LLLTSHFLGESLG/GG

&lt;400&gt; 167

```

taaagtccctg tgtatgacat gacatagtat ttgcgtaatt taa atg tac ata aag      55
                                Met Tyr Ile Lys
                                -35
atg gag tct gtc acc ttg tca cca gcc cca gtc ttc ccc gtc cct gca      103
Met Glu Ser Val Thr Leu Ser Pro Ala Pro Val Phe Pro Val Pro Ala
      -30                      -25                      -20

```

101

```

car ctc ctt tta ctg aca tcc cat ttt cta ggc gag tcc ctt ggt gga      151
Gln Leu Leu Leu Leu Thr Ser His Phe Leu Gly Glu Ser Leu Gly Gly
-15                      -10                      -5                      1
ggc aca ctg ctt gtc cca ctc ctc ccc cca ggg                        184
Gly Thr Leu Leu Val Pro Leu Leu Pro Pro Gly
                    5                      10

```

&lt;210&gt; 168

&lt;211&gt; 218

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 97..216

&lt;221&gt; sig\_peptide

&lt;222&gt; 97..177

&lt;223&gt; Von Heijne matrix

score 7.40000009536743

seq ILLLTICAAGIXG/TR

&lt;400&gt; 168

```

ccttcctccc ggcacaggc tgccggtcca ccgcttgcta atggcagccg gggtctccct      60
gggacagcaa gacctccgct caggcccttc ttccga atg ckc cam gcm ctc ctg      114
                                   Met Xaa Xaa Ala Leu Leu
                                   -25
cga tct aga atg att cag ggc agg atc ctg ctc ctg acc atc tgc gct      162
Arg Ser Arg Met Ile Gln Gly Arg Ile Leu Leu Leu Thr Ile Cys Ala
-20                      -15                      -10
gcc ggc att rgt ggg act cgt cag ttt ggc tat aac ctc tct atc atc      210
Ala Gly Ile Xaa Gly Thr Arg Gln Phe Gly Tyr Asn Leu Ser Ile Ile
-5                      1                      5                      10
aat gac cc                        218
Asn Asp

```

&lt;210&gt; 169

&lt;211&gt; 480

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 317..478

&lt;221&gt; sig\_peptide

&lt;222&gt; 317..457

&lt;223&gt; Von Heijne matrix

score 7.40000009536743

seq SCLFSXAWLXCXC/HG

&lt;400&gt; 169

```

gtctcgtggg ctggtcccca ggggtccct cccgaacag ctgctgctcc agggaggaag      60
cggcgyrrgt gmtgtccagc ttcccgggtgc tgaaaaccgg agggctcgtc atccaccact      120
accatgtaag ggccatgaga agggctcatc ctggcgcasg cggacatgga ggaggactta      180
ttccagctaa ggcagctgcc gggtgtgaaa ttccgtcgca caggcgagag tgcaagggtca      240
gaggacgaca cggcttcagg agagcatgaa gtccagattg aaggggtcca cgtgggccta      300
gaggctgtgg agctgg atg atg ggg cak ctg tgc cca agg agt ttg cca atc      352
                                   Met Met Gly Xaa Leu Cys Pro Arg Ser Leu Pro Ile
                                   -45                      -40
cca ccg atg ata ctt tca tgg tgg aag atg cag tgg aag cca ttg gct      400
Pro Pro Met Ile Leu Ser Trp Trp Lys Met Gln Trp Lys Pro Leu Ala
-35                      -30                      -25                      -20

```

102

```

ttg gaa aat ttc agt gga agc tgt ctg ttc tca mtg gct tgg ctt kga      448
Leu Glu Asn Phe Ser Gly Ser Cys Leu Phe Ser Xaa Ala Trp Leu Xaa
               -15                      -10                      -5
tgc tsa tgc cat gga gat gat gat ctc agc at                        480
Cys Xaa Cys His Gly Asp Asp Asp Leu Ser
              1                      5

```

&lt;210&gt; 170

&lt;211&gt; 280

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 135..278

&lt;221&gt; sig\_peptide

&lt;222&gt; 135..179

&lt;223&gt; Von Heijne matrix

score 7.40000009536743

seq LLQLLAFSFLGNS/VE

&lt;221&gt; misc\_feature

&lt;222&gt; 104

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 170

```

ttcttttgggc tcgggggctc ccggagcagg gcgagagctc gcgtcgccgg aaaggaagac      60
gggaagaaag ggcaggcggc tcggcgggcg tcttctccac tccntgccgc gcccgtggc      120
tgcaggggagc cggc atg ggg ctt ctc cag ttg cta gct ttc agt ttc tta      170
               Met Gly Leu Leu Gln Leu Leu Ala Phe Ser Phe Leu
               -15                      -10                      -5
ggt aat tcc gtg gaa acg gtg cgg gga ggc gga cgg act tgg gca tgg      218
Gly Asn Ser Val Glu Thr Val Arg Gly Gly Gly Arg Thr Trp Ala Trp
              1                      5                      10
gga agg aaa acc caa aag ctg ctt gct cac ctt cgt ggg atc ctg ggg      266
Gly Arg Lys Thr Gln Lys Leu Leu Ala His Leu Arg Gly Ile Leu Gly
              15                      20                      25
gct tgg gas agg ga                        280
Ala Trp Xaa Arg
30

```

&lt;210&gt; 171

&lt;211&gt; 103

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 28..102

&lt;221&gt; sig\_peptide

&lt;222&gt; 28..69

&lt;223&gt; Von Heijne matrix

score 7.40000009536743

seq LVLVHSSLSKTLK/QK

&lt;400&gt; 171

```

actgggatgc agaggctgca gtgagcc atg ttg gtg ctg gtg cac tcc agc ctg      54
               Met Leu Val Leu Val His Ser Ser Leu
               -10
agc aag acc ttg tct cag aaa aaa aaa aag ttc aca aas ccc acc agg g      103

```



Ser Lys Thr Leu Ser Gln Lys Lys Lys Lys Phe Thr Xaa Pro Thr Arg  
 -5 1 5 10

<210> 172  
 <211> 218  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 73..216

<221> sig\_peptide  
 <222> 73..129  
 <223> Von Heijne matrix  
 score 7.40000009536743  
 seq LILVISCLLLAFE/CV

<400> 172  
 caattttgtt gatcttttca aaaaaccagc tcctggattc attaatTTTT tgaagggttt 60  
 tttgatgtct ct atg tcc ttc agt tct gct ctg att tta gtt att tct tgc 111  
 Met Ser Phe Ser Ser Ala Leu Ile Leu Val Ile Ser Cys  
 -15 -10  
 ctt ctg cta gct ttt gaa tgt gtt tgc tct tgc ttt tct ggt tct ttt 159  
 Leu Leu Leu Ala Phe Glu Cys Val Cys Ser Cys Phe Ser Gly Ser Phe  
 -5 1 5 10  
 aat tgt gat gtt agg gtg tca att tgc gat ctt tcc tgc ttt ctc ttg 207  
 Asn Cys Asp Val Arg Val Ser Ile Ser Asp Leu Ser Cys Phe Leu Leu  
 15 20 25  
 tgg ggc aag gg 218  
 Trp Gly Lys

<210> 173  
 <211> 380  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 295..378

<221> sig\_peptide  
 <222> 295..360  
 <223> Von Heijne matrix  
 score 7.40000009536743  
 seq CLXVFLLTDRTLS/CR

<400> 173  
 tattggttat tctagttata cattagtcta aatttttttc aaagttttca acttctttgc 60  
 ctttggtttg aatttctctc tgtagcttgg agtagtttga tcatctgaag ctttcttctc 120  
 tcaactcatc aaagtcattc tccatccagc ttgtgtccat tgctgggtgag gaactgtgtt 180  
 ccttcggagg aggagaggtg ctctgctttt ttgagtttcc agtttttctg ctctgttttt 240  
 tccccatctt tgtggtttta tctacttttg gtctttgatg ctggtgatgt acag atg 297  
 Met  
 ggt ttt tgg tgt gga tgt cct ttc tgt ttg twa gtt ttc ctt cta aca 345  
 Gly Phe Trp Cys Gly Cys Pro Phe Cys Leu Xaa Val Phe Leu Leu Thr  
 -20 -15 -10  
 gac agg acc ctc agc tgc agg tct gtt gga gtt gc 380  
 Asp Arg Thr Leu Ser Cys Arg Ser Val Gly Val  
 -5 1 5

<210> 174  
 <211> 139

<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 59..139

<221> sig\_peptide  
<222> 59..103  
<223> Von Heijne matrix  
score 7.30000019073486  
seq LLSLSLWGISTLS/ST

<400> 174  
ataacagaat gatttacatt cctttgggta tatacccagt gatgggatat atgtgtca 58  
atg gta tta ctg tct tta agt ctt tgg ggc atc tcc aca ctg tct tcc 106  
Met Val Leu Leu Ser Leu Ser Leu Trp Gly Ile Ser Thr Leu Ser Ser  
-15 -10 -5 1  
aca aca att gaa cta att tac acc ccc atc ggg 139  
Thr Thr Ile Glu Leu Ile Tyr Thr Pro Ile Gly  
5 10

<210> 175  
<211> 122  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 38..121

<221> sig\_peptide  
<222> 38..112  
<223> Von Heijne matrix  
score 7.30000019073486  
seq LLHVHVSFLPPVFS/TQ

<400> 175  
ctacctgtcc ttgcgcacca cccttgtctg ggccttc atg gcc tct ctg ctg agt 55  
Met Ala Ser Leu Leu Ser  
-25 -20  
ggc ttt act agc ttc tgt ctt ttg cac gtt cac tct ttc ctg cct cca 103  
Gly Phe Thr Ser Phe Cys Leu Leu His Val His Ser Phe Leu Pro Pro  
-15 -10 -5  
gtg ttt tcc acc cag aat g 122  
Val Phe Ser Thr Gln Asn  
1

<210> 176  
<211> 300  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 175..300

<221> sig\_peptide  
<222> 175..264  
<223> Von Heijne matrix  
score 7.30000019073486  
seq AILLXXWEAGSEA/VR

<221> misc\_feature  
 <222> 51..52,63,239  
 <223> n=a, g, c or t

<400> 176  
 aaaaactcta aaagaaggac gcatttttagg taagatctag tggctagatc nncaggggtgg 60  
 gcncggttct tgtggaaatc agtcaagaaa gatcggattc gcggttattt atgcaaatca 120  
 tctgggtgga ttgtgtacgg agttaaactg cgccttctgg accgggtctg aaca atg 177  
 Met  
 -30  
 gag act gcg cta saa tka acg cca cag aaa agg caa gtt atg ttt ctt 225  
 Glu Thr Ala Leu Xaa Xaa Thr Pro Gln Lys Arg Gln Val Met Phe Leu  
 -25 -20 -15  
 gct ata ttg ttg cnt twg tgg gag gct ggc tct gag gca gth agg tat 273  
 Ala Ile Leu Leu Xaa Xaa Trp Glu Ala Gly Ser Glu Ala Val Arg Tyr  
 -10 -5 1  
 tcc ata cca gaa gaa aca gaa agt ggc 300  
 Ser Ile Pro Glu Glu Thr Glu Ser Gly  
 5 10

<210> 177  
 <211> 466  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 268..465

<221> sig\_peptide  
 <222> 268..372  
 <223> Von Heijne matrix  
 score 7.30000019073486  
 seq LDLLGSSSPPTSA/SQ

<400> 177  
 cttaaacttt attatgttgk kttcaciaaag agcagccttt gttgactttg aaatcattgc 60  
 ttcagtattc tagaaaaatct tgtttttgtt aaacatgggc agtaacttac tatttttgta 120  
 tagttgttgt wcatckttacc cccaccctgt tttaaaaaata aaaagtagtt gtcagattac 180  
 tttggcttta gaagtacctt ttcacttgcc ttagaatctt cattactttg agcctacact 240  
 ccacctctta ttggaacttc atgaaga atg atg ttg gat ttc gct ctg tcg ccc 294  
 Met Met Leu Asp Phe Ala Leu Ser Pro  
 -35 -30  
 agg cta gag cgc agt ggt ctg atc atg gct tgc tgt acc ctt gac ctc 342  
 Arg Leu Glu Arg Ser Gly Leu Ile Met Ala Cys Cys Thr Leu Asp Leu  
 -25 -20 -15  
 ctg ggt tca agc agt cct ccc acc tca gcc tcc cag gtg gct ggg act 390  
 Leu Gly Ser Ser Ser Pro Pro Thr Ser Ala Ser Gln Val Ala Gly Thr  
 -10 -5 1 5  
 ggg cat gtg cca cca cac cca gct agt ttt ttt tac ttt ktt gta wga 438  
 Gly His Val Pro Pro His Pro Ala Ser Phe Phe Tyr Phe Xaa Val Xaa  
 10 15 20  
 cag gtc tac tat gtt tcg cag ctg atc t 466  
 Gln Val Tyr Tyr Val Ser Gln Leu Ile  
 25 30

<210> 178  
 <211> 222  
 <212> DNA  
 <213> Homo sapiens

<220>

<221> CDS  
<222> 30..221

<221> sig\_peptide  
<222> 30..95  
<223> Von Heijne matrix  
score 7.19999980926514  
seq. QVFFLVFPDGVPR/QP

<400> 178  
acgtcggacc cggaggccct gaatgcccc atg cgc acc cca cag ctc gcg ctc 53  
Met Arg Thr Pro Gln Leu Ala Leu  
-20 -15  
ctg caa gtg ttc ttt ctg gtg ttc ccc gat ggc gtc cgg cct cag ccc 101  
Leu Gln Val Phe Phe Leu Val Phe Pro Asp Gly Val Arg Pro Gln Pro  
-10 -5 1  
tct tcc tcc cca tca ggg gca gtg ccc acg tct ttg gag ctg cag cga 149  
Ser Ser Ser Pro Ser Gly Ala Val Pro Thr Ser Leu Glu Leu Gln Arg  
5 10 15  
ggg acg gat ggc gga acc ctc cag tcc cct tca gag gcg act gca act 197  
Gly Thr Asp Gly Gly Thr Leu Gln Ser Pro Ser Glu Ala Thr Ala Thr  
20 25 30  
cgc ccg gcc gtg ccc gga ctc cgg g 222  
Arg Pro Ala Val Pro Gly Leu Arg  
35 40

<210> 179  
<211> 171  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 33..170  
  
<221> sig\_peptide  
<222> 33..95  
<223> Von Heijne matrix  
score 7.19999980926514  
seq SWPLLAASGLRG/LE

<400> 179  
ccttttgcct tcaaccttcg agccgccacg ta atg cca cgt ccc cgc gca tgc 53  
Met Pro Arg Pro Arg Ala Cys  
-20 -15  
gca tct tgg ccg ctg ctg gcg gct gtt tcc ggg ctt aga ggg ctg gag 101  
Ala Ser Trp Pro Leu Leu Ala Ala Val Ser Gly Leu Arg Gly Leu Glu  
-10 -5 1  
tgg ccg ccg agt tgg agg cgg gtg gtg gca gca gta gga gtg tgt aga 149  
Trp Pro Pro Ser Trp Arg Arg Val Val Ala Ala Val Gly Val Cys Arg  
5 10 15  
gtg cgg gat tgg ggg ccc cgg g 171  
Val Arg Asp Trp Gly Pro Arg  
20 25

<210> 180  
<211> 245  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 177..245

```

<221> sig_peptide
<222> 177..227
<223> Von Heijne matrix
      score 7.19999980926514
      seq FLLLLISVLTVIWF/WK

<400> 180
tgtaatttttc cttgccaaaa agcttagttt catctttttat aaatataccta taatgccaaag      60
ttgattgcat ggtcagagtg aatctgtgct gtaccawat tcagtagcct tctcctatcc      120
aacaaagtgt tttgtaaata ggaggtaaat gaatgagtgg atggatggag ggatga atg      179
                                         Met
aat gga att ttc ttg ctc ttg atc tct gtc tta aca gtg att tgg ttt      227
Asn Gly Ile Phe Leu Leu Leu Ile Ser Val Leu Thr Val Ile Trp Phe
   -15                               -10                               -5
tgg aag aca cac ccg ggg      245
Trp Lys Thr His Pro Gly
1                               5

<210> 181
<211> 241
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 160..240

<221> sig_peptide
<222> 160..213
<223> Von Heijne matrix
      score 7.19999980926514
      seq XLLCIIXLYLIRG/SE

<400> 181
gttgactttt ctctctgctg aggcagaaaa atgcttccat agtccatgca gcaatgttta      60
aaacaaggga tttcggtccc ccctcvcctt ttgtgtaggc tggtaataa actctgtggt      120
tywtacgatt gtcgtgaawa ttcagagtgc tccctgcga atg gtt ttc cta gta      174
                                         Met Val Phe Leu Val
                                         -15
kct ctg ttg tgt atc att kct ctt tat ttg att cgt ggt tct gag tgg      222
Xaa Leu Leu Cys Ile Ile Xaa Leu Tyr Leu Ile Arg Gly Ser Glu Trp
   -10                               -5                               1
amc cta cca ccg aac tgg g      241
Xaa Leu Pro Pro Asn Trp
5

<210> 182
<211> 263
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 103..261

<221> sig_peptide
<222> 103..156
<223> Von Heijne matrix
      score 7.19999980926514
      seq LFFLLRIALASWA/LF

<400> 182

```

108

```

gggttatcta acctgttcca ttgttcctg tatcagtttc tgtaccgata ccatgctggt 60
ttggttactg tagtcttgta gtatagttta aagtcagata gc atg atg act cta 114
                               Met Met Thr Leu
                               -15
gct ttg ttc ttt ttg ctt agg att gct ttg gct agt tgg gct ctc ttt 162
Ala Leu Phe Phe Leu Leu Arg Ile Ala Leu Ala Ser Trp Ala Leu Phe
                               -10      -5      1
tgg atc cat atg aat ttt aga aga gct ttt ttc cac tta cgg tgg ttt 210
Trp Ile His Met Asn Phe Arg Arg Ala Phe Phe His Leu Arg Trp Phe
                               5      10      15
gat atc aat agc act gaa tct gta aat tgc ttt ggg cag tat ggc cta 258
Asp Ile Asn Ser Thr Glu Ser Val Asn Cys Phe Gly Gln Tyr Gly Leu
                               20      25      30
gcg gg 263
Ala
35

```

<210> 183  
 <211> 170  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 60..170

<221> sig\_peptide  
 <222> 60..146  
 <223> Von Heijne matrix  
 score 7.09999990463257  
 seq SLLVFCLNDLSNA/VX

```

<400> 183
ttccatgtgg agatgrraag aatatatatt ctgtggttat tgggtagagt gttctatag 59
atg tct att agg tct aat tgg tct agt gtc gaa tct aag tct aga att 107
Met Ser Ile Arg Ser Asn Trp Ser Ser Val Glu Ser Lys Ser Arg Ile
                               -25      -20      -15
tct tta tta gtt ttc tgc ctc aat gat ctw tck aat gcw gtc arg wgg 155
Ser Leu Leu Val Phe Cys Leu Asn Asp Leu Ser Asn Ala Val Xaa Xaa
                               -10      -5      1
ggm att gaa rtc ccc 170
Gly Ile Glu Xaa Pro
5

```

<210> 184  
 <211> 443  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 83..442

<221> sig\_peptide  
 <222> 83..130  
 <223> Von Heijne matrix  
 score 7.09999990463257  
 seq IPLFLGVLAYCTG/SV

```

<400> 184
ctttccagca aggggataag agaggcctgg aagaacctgc ccagcctggg cctcaggaag 60
cagcatcgga ggtgcctcag cc atg gca tgg atc cct ctc ttc ctc ggc gtc 112
                               Met Ala Trp Ile Pro Leu Phe Leu Gly Val

```

																-15						-10						
ctt	gct	tac	tgc	aca	gga	tcc	gtg	gcc	tcc	tat	gag	ctg	act	cac	cca	160												
Leu	Ala	Tyr	Cys	Thr	Gly	Ser	Val	Ala	Ser	Tyr	Glu	Leu	Thr	His	Pro													
						-5						5						10										
ccc	tca	gtg	tcc	gtg	tcc	cca	gga	cag	aca	gcc	agc	atc	acc	tgc	tct	208												
Pro	Ser	Val	Ser	Val	Ser	Pro	Gly	Gln	Thr	Ala	Ser	Ile	Thr	Cys	Ser													
						15						20						25										
gga	gat	aaa	ttg	ggg	gat	aaa	tat	gct	tgc	tgg	tat	cag	cag	aag	cca	256												
Gly	Asp	Lys	Leu	Gly	Asp	Lys	Tyr	Ala	Cys	Trp	Tyr	Gln	Gln	Lys	Pro													
						30						35						40										
ggc	cag	tcc	cct	gtg	ctg	gtc	atc	tat	caa	gat	agc	aag	cgg	ccc	tca	304												
Gly	Gln	Ser	Pro	Val	Leu	Val	Ile	Tyr	Gln	Asp	Ser	Lys	Arg	Pro	Ser													
						45						50						55										
ggg	atc	cct	gag	cga	ttc	tct	ggc	tcc	aac	tct	ggg	aac	aca	gcc	act	352												
Gly	Ile	Pro	Glu	Arg	Phe	Ser	Gly	Ser	Asn	Ser	Gly	Asn	Thr	Ala	Thr													
						60						65						70										
ctg	acc	atc	agc	ggg	acc	cag	gct	atg	gat	gag	gct	gac	tat	tac	tgt	400												
Leu	Thr	Ile	Ser	Gly	Thr	Gln	Ala	Met	Asp	Glu	Ala	Asp	Tyr	Tyr	Cys													
						75						80						85										
cag	gcg	tgg	gac	agc	agc	act	gtg	gta	ttc	ggc	gga	ggg	acc	a		443												
Gln	Ala	Trp	Asp	Ser	Ser	Thr	Val	Val	Phe	Gly	Gly	Gly	Thr															
						95						100																

```
<210> 185
<211> 427
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 332..427
```

```
<221> sig_peptide
<222> 332..418
<223> Von Heijne matrix
      score 7.09999990463257
      seq FCFXLCFGRSSLG/CR
```

<400>	185								
taagtttata	yhtctgaatc	tgaaatcaga	atatatatat	ttaatttttc	aatttttaaaa				60
atgttaccct	gtgtgagaca	aaacaaaaca	gtgactagaa	ccctccttgt	gggctaaatt				120
tgagtttgct	tcttcataat	gtttttaaatg	cttcacaaac	atttttcttt	ggtatattga				180
gcaaaatgaa	ttgaagtata	tttactgagt	gatgattatt	gaggaaaaac	tcaaagatct				240
gctgtaagca	ctagagttga	aggactagcc	caacagctcc	tcaggcacct	ttgggtatat				300
tgagttgccc	cccctgactt	tgaacacatc	t atg gtc	tgt gtc atc	ttc aaa				352
			Met Val	Cys Val	Ile Phe Lys				

																-25	
gag	ctc	atg	gaa	ttt	gaa	ttc	cct	ggg	ttt	tgt	ttt	tgh	ctt	tgt	ttt	400	
Glu	Leu	Met	Glu	Phe	Glu	Phe	Pro	Gly	Phe	Cys	Phe	Xaa	Leu	Cys	Phe		
				-20					-15					-10			
gga	cgg	agc	tcg	ctc	tgt	tgc	cga	rac								427	
Gly	Arg	Ser	Ser	Leu	Cys	Cys	Arg	Xaa									
		-5			1												

```
<210> 186
<211> 365
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 130..363
```

110

```

<221> sig_peptide
<222> 130..219
<223> Von Heijne matrix
      score 7.09999990463257
      seq SCLALXTLAVVYA/AL

<400> 186
aacgagtcctt tgggaacgtg gtccacccag ggatgtaaaa ctgtgcttac cgatgcatcc      60
catacgaat gcttatgtga tcgtctctct accttcgcca ttttggctca gcaacctaga      120
gaaataatc atg gaa tcc tct ggc aca cct tca gtt acc cta ata gta ggc      171
      Met Glu Ser Ser Gly Thr Pro Ser Val Thr Leu Ile Val Gly
      -30          -25          -20
agt ggt ctt tct tgc ttg gcc ttg atb acc cta gca gtt gtc tat gca      219
Ser Gly Leu Ser Cys Leu Ala Leu Xaa Thr Leu Ala Val Val Tyr Ala
      -15          -10          -5
gca tta tgg mgg tac ata cgc tct gag aga tcc ata ata cta att aac      267
Ala Leu Trp Arg Tyr Ile Arg Ser Glu Arg Ser Ile Ile Leu Ile Asn
      1          5          10          15
ttc tgc ctg tct atc atc tca tcc aat atc ctc ata ctg gtt gga cag      315
Phe Cys Leu Ser Ile Ile Ser Ser Asn Ile Leu Ile Leu Val Gly Gln
      20          25          30
act cag aca cat aat aaa gag tat ctg cac aac cac cac tgc att ttt      363
Thr Gln Thr His Asn Lys Glu Tyr Leu His Asn His His Cys Ile Phe
      35          40          45
gc      365

<210> 187
<211> 260
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 86..259

<221> sig_peptide
<222> 86..178
<223> Von Heijne matrix
      score 7.09999990463257
      seq LXFLASSFCFGEA/DS

<221> misc_feature
<222> 143
<223> n=a, g, c or t

<400> 187
ttttggaaca gggtaggcat tttgtttatt gtttgcttgc ttctaggtgt tttcgccatc      60
agggtgtatt ggaggctgac actta atg ggt gtg tgt tgc gcc cag aac tgc      112
      Met Gly Val Cys Cys Ala Gln Asn Cys
      -30          -25
tcg gtg tcg ggg ktc waa agr aat gcg ctg ntg ttc ttg gct tca agt      160
Ser Val Ser Gly Xaa Xaa Arg Asn Ala Leu Xaa Phe Leu Ala Ser Ser
      -20          -15          -10
ttc tgc ttt gga gaa gca gat tca gga agt agg tgt tgc tta aaa ata      208
Phe Cys Phe Gly Glu Ala Asp Ser Gly Ser Arg Cys Cys Leu Lys Ile
      -5          1          5          10
att ctt ggt ttt tat cta atc aga tat tca ttg att acc tac cag gtg      256
Ile Leu Gly Phe Tyr Leu Ile Arg Tyr Ser Leu Ile Thr Tyr Gln Val
      15          20          25
cgt g      260
Arg

```



<210> 188  
 <211> 172  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 52..171

<221> sig\_peptide  
 <222> 52..105  
 <223> Von Heijne matrix  
 score 7.09999990463257  
 seq LFFFLKWSHPGWS/AT

<221> misc\_feature  
 <222> 112  
 <223> n=a, g, c or t

<400> 188  
 ttaatggaat atattatagt acactagcat gctggaaaga atgaaaataa t atg aaa 57  
 Met Lys  
 att ctt tac ctt ttt ttc ttt ttg aaa tgg agt cac cca ggc tgg agt 105  
 Ile Leu Tyr Leu Phe Phe Phe Leu Lys Trp Ser His Pro Gly Trp Ser  
 -15 -10 -5  
 gca acg ncg tgg tct tgg cac act gca acc tcc gcc tcc ctg att caa 153  
 Ala Thr Xaa Trp Ser Trp His Thr Ala Thr Ser Ala Ser Leu Ile Gln  
 1 5 10 15  
 gtg att ctc ccg cct tgg g 172  
 Val Ile Leu Pro Pro Trp  
 20

<210> 189  
 <211> 150  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 47..148

<221> sig\_peptide  
 <222> 47..124  
 <223> Von Heijne matrix  
 score 7.09999990463257  
 seq LFLSGCFLFLSXC/XI

<400> 189  
 tatcacwtct aagagatttc tggatgaaact tgtggatttt ctatac atg aca cca 55  
 Met Thr Pro  
 -25  
 tgt ttt ctg caa atg gac aat ttg act cct ctt ttc cta tct gga tgc 103  
 Cys Phe Leu Gln Met Asp Asn Leu Thr Pro Leu Phe Leu Ser Gly Cys  
 -20 -15 -10  
 ttt tta ttt ctc tct cwt tgc wtg att tat ttg gct agg att ttg gg 150  
 Phe Leu Phe Leu Ser Xaa Cys Xaa Ile Tyr Leu Ala Arg Ile Leu  
 -5 1 5

<210> 190  
 <211> 339  
 <212> DNA  
 <213> Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 195..338

&lt;221&gt; sig\_peptide

&lt;222&gt; 195..314

&lt;223&gt; Von Heijne matrix

score 7

seq ITCKLCLCEQSG/QD

&lt;400&gt; 190

agtcttgcaa agtgtaaagc tgtcagccgc agagcacgga ggaaagacgg agagaatgga 60

agagctcctg tccggtgtgc cagcagcccg gactggcggg gagcgcgagg gaggtackg 120

agaagcccg cgacggagga acgcaggtct gctgccaggg attgaggaga ctgaagaacg 180

ctgaagacag gctg atg ggc tca gct ggt agg ctc cac tat ctc gsc atg 230

Met Gly Ser Ala Gly Arg Leu His Tyr Leu Xaa Met

-40

-35

-30

act gct gaa aat ccc act cct gga gac ctg gct ccg kcc ccc ctc atc 278

Thr Ala Glu Asn Pro Thr Pro Gly Asp Leu Ala Pro Xaa Pro Leu Ile

-25

-20

-15

act tgc aaa ctc tgc ctg tgt gag cag tct crt gga caa gat gac cac 326

Thr Cys Lys Leu Cys Leu Cys Glu Gln Ser Xaa Gly Gln Asp Asp His

-10

-5

1

act cca gga atg c 339

Thr Pro Gly Met

5

&lt;210&gt; 191

&lt;211&gt; 359

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 96..359

&lt;221&gt; sig\_peptide

&lt;222&gt; 96..242

&lt;223&gt; Von Heijne matrix

score 7

seq VIVLLSAAPCLLS/CF

&lt;221&gt; misc\_feature

&lt;222&gt; 340

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 191

tacaagagtt tttgctgaaa gttttaagtt gataagatgc agagaattgg gggaatgtat 60

aataaatcag gtttcattgt tatattatcc accac atg aat cac ctt cct cct 113

Met Asn His Leu Pro Pro

-45

aac cat tat agg mgc cat gtg ttc aca tgt cat gtg gac cag tat tta 161

Asn His Tyr Arg Xaa His Val Phe Thr Cys His Val Asp Gln Tyr Leu

-40

-35

-30

act gtg gaa acc gcg ggt ggc atg gag aag gag gca gtg tcc gtg act 209

Thr Val Glu Thr Ala Gly Gly Met Glu Lys Glu Ala Val Ser Val Thr

-25

-20

-15

gtg ctg ctc tcc gca gcc ccc tgc ctg ctg tcc tgt ttc ctc ggc tcc 257

Val Leu Leu Ser Ala Ala Pro Cys Leu Leu Ser Cys Phe Leu Gly Ser

-10

-5

1

5

tcg gtg tct gga ctg gcg ttc tgg gtt tcc cag cag aaa act aaa ggg 305

```

Ser Val Ser Gly Leu Ala Phe Trp Val Ser Gln Gln Lys Thr Lys Gly
      10      15      20
cca gag agg tgt aaa aac aca cac cac tbg gca gnt aat aat ttc ccc      353
Pro Glu Arg Cys Lys Asn Thr His His Xaa Ala Xaa Asn Asn Phe Pro
      25      30      35

gcg agg
Ala Arg

<210> 192
<211> 264
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 138..263

<221> sig_peptide
<222> 138..257
<223> Von Heijne matrix
      score 7
      seq FLFMLPLWCSIGT/CT

<400> 192
ttgagcttaa ggccagggtat atggggtcac acttgtaatc tcagtgtctt gggagggtga      60
gggaaaagga tagcttgagt ccaggagttc gagatcatcc tgggcaacat agcaagatcc      120
tgtctctaca aaaccta atg aac aaa att aaa gaa aac aca cac aca cac      170
          Met Asn Lys Ile Lys Glu Asn Thr His Thr His
          -40          -35          -30
aca cac aca cac aca cac aaa aac aac acc aaa cta gtg tca aac cta      218
Thr His Thr His Thr His Lys Asn Asn Thr Lys Leu Val Ser Asn Leu
          -25          -20          -15
ttc ctt ttt atg tta cct ctc tgg tgc tcc att ggc act tgc aca g      264
Phe Leu Phe Met Leu Pro Leu Trp Cys Ser Ile Gly Thr Cys Thr
          -10          -5          1

<210> 193
<211> 301
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 147..299

<221> sig_peptide
<222> 147..272
<223> Von Heijne matrix
      score 7
      seq LFLYSLFTENVLA/HP

<400> 193
tgtattgttt mmmttattta ctagtatgca gatctgggtt tcattctttt catattgaat      60
ttcgttatgg gtagaatcat ttgcaaacat ttctagacat ttttaaagat ctatttaatt      120
tgtttaagaa tggaaaacat aaaata atg cat gat tct tca ggc aag aat aat      173
          Met His Asp Ser Ser Gly Lys Asn Asn
          -40          -35
ttc aga aag ata cct gtt gta aat tta att tat ctc tat gta gac ata      221
Phe Arg Lys Ile Pro Val Val Asn Leu Ile Tyr Leu Tyr Val Asp Ile
          -30          -25          -20
cat ata cat aaa tta ttt tta tat agt ctc ttt aca gaa aat gta ttg      269
His Ile His Lys Leu Phe Leu Tyr Ser Leu Phe Thr Glu Asn Val Leu
          -15          -10          -5

```

114

gca cat cct tgc att gtt cta cgc cgc cta tg  
 Ala His Pro Cys Ile Val Leu Arg Arg Leu  
 1 5

301

<210> 194  
 <211> 215  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 105..215

<221> sig\_peptide  
 <222> 105..203  
 <223> Von Heijne matrix  
 score 7  
 seq LFFLFVGPFSCLG/SY

<400> 194  
 gctctgactg cagcctccca gggaatgcgc ggccgagggga atgcgcgcag tcacaggccc 60  
 tgggagttag ctggtgcccgc gcgacctggc acccgcgccct ggat atg ggg cgt cta 116  
 Met Gly Arg Leu  
 -30  
 cat cgt ccc agg agc agc acc agc tac agg aac ctg ccg cat ctg ttt 164  
 His Arg Pro Arg Ser Ser Thr Ser Tyr Arg Asn Leu Pro His Leu Phe  
 -25 -20 -15  
 ctg ttt ttc ctc ttc gtg gga ccc ttc agc tgc ctc ggg agt tac agc 212  
 Leu Phe Phe Leu Phe Val Gly Pro Phe Ser Cys Leu Gly Ser Tyr Ser  
 -10 -5 1  
 cgg 215  
 Arg

<210> 195  
 <211> 209  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 78..209

<221> sig\_peptide  
 <222> 78..158  
 <223> Von Heijne matrix  
 score 7  
 seq RLLLLLLXLPLP/PP

<221> misc\_feature  
 <222> 73..74  
 <223> n=a, g, c or t

<400> 195  
 tcattcactg attagatcca gcgctgagag gcagcactgc tccttctctc acgccaactg 60  
 agtctcttga tcnntac atg caa tcc cag gca gct cgc gaa cac aaa ccc 110  
 Met Gln Ser Gln Ala Ala Arg Glu His Lys Pro  
 -25 -20  
 ggg ghc agc cgc cta ctg ctg ctg ctg ctg cwg ctg ccg ctg cct 158  
 Gly Xaa Ser Arg Leu Leu Leu Leu Leu Leu Xaa Leu Pro Leu Pro  
 -15 -10 -5  
 ccg ccg gkv ctg cga acc cgg gdy ttt tca wgc acc aca ctc acc gcm 206  
 Pro Pro Xaa Leu Arg Thr Arg Xaa Phe Ser Xaa Thr Thr Leu Thr Ala

```

1          5          10          15          209
ggg
Gly

<210> 196
<211> 363
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 287..361

<221> sig_peptide
<222> 287..331
<223> Von Heijne matrix
      score 7
      seq LWSLACLSPPAVQ/LG

<400> 196
ttacattgta atatataaat aattatacaa ctcaccataa cgtagaatca gtgggagccc      60
tgagcttggt ttcctgcaac tagatggtcc caactagacc aggtgatggg agacaatgac      120
agatcattag gcattagatt atcataagga gcatacaacc tagatccctt gcatgtgcag      180
ttaataatag gttttgcact tctatgagga tctaattgcgg cctctgatct gacaaggggc      240
ggastcaggc agtaatggga gcaatgggga gcggttttca atacag atg agg ctt      295
                                   Met Arg Leu
                                   -15
tgg tca ctt gcc tgc ctt tca cct cct gct gtg cag ctt ggt tcc caa      343
Trp Ser Leu Ala Cys Leu Ser Pro Pro Ala Val Gln Leu Gly Ser Gln
      -10          -5          1
cag gcc acg gac tgg tgg tc      363
Gln Ala Thr Asp Trp Trp
5          10

<210> 197
<211> 155
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 58..153

<221> sig_peptide
<222> 58..132
<223> Von Heijne matrix
      score 7
      seq IFSFFFFITLVRG/SI

<400> 197
tagtggttatt catagtagta tctgaagacc ttttgtattc ttgtgggatc agttgta      57
atg tca cct ttg ttt att ctg att gtg ctt att tgg atc ttc tct ttc      105
Met Ser Pro Leu Phe Ile Leu Ile Val Leu Ile Trp Ile Phe Ser Phe
-25          -20          -15          -10
ttt ttc ttt att act cta gtt agg ggg tct atc aat ctt ttt ttt ttt      153
Phe Phe Phe Ile Thr Leu Val Arg Gly Ser Ile Asn Leu Phe Phe Phe
      -5          1          5
tt      155

<210> 198
<211> 135
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> CDS
<222> 60..134

<221> sig_peptide
<222> 60..125
<223> Von Heijne matrix
      score 7
      seq STFLFFLFFSVFC/FF

<400> 198
ttgcctctta aaaggccaca cttcttaata ctatcaaatt ggctattaag tttcaacaa      59
atg aat ttg ggg gga cat tca gat cat agc act ttt ctt ttc ttt ctt      107
Met Asn Leu Gly Gly His Ser Asp His Ser Thr Phe Leu Phe Phe Leu
      -20              -15              -10
ttt ttt tct gtt ttt tgt ttt ttt ttt t      135
Phe Phe Ser Val Phe Cys Phe Phe Phe
      -5              1

<210> 199
<211> 320
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 46..318

<221> sig_peptide
<222> 46..108
<223> Von Heijne matrix
      score 6.90000009536743
      seq VTFLALVGAVLY/LY

<221> misc_feature
<222> 9
<223> n=a, g, c or t

<400> 199
gctggagcng ccgatccgag acgtggcthc ctgggcggca gaacc atg ttg gac ttc      57
                               Met Leu Asp Phe
                               -20
gcg atc ttc gcc gtt acc ttc ttg ctg gcg ttg gtg gga gcc gtg ctc      105
Ala Ile Phe Ala Val Thr Phe Leu Leu Ala Leu Val Gly Ala Val Leu
      -15              -10              -5
tac ctc tat ccg gct tcc aga caa gct gca gga att cca ggg att act      153
Tyr Leu Tyr Pro Ala Ser Arg Gln Ala Ala Gly Ile Pro Gly Ile Thr
      1              5              10              15
cca act gaa gaa aaa gat ggt aat ctt cca gat att gtg aat agt gga      201
Pro Thr Glu Glu Lys Asp Gly Asn Leu Pro Asp Ile Val Asn Ser Gly
      20              25              30
agt ttg cat gag tbc ctg gtt aat ttg cat gag aga tat ggg cct gtg      249
Ser Leu His Glu Xaa Leu Val Asn Leu His Glu Arg Tyr Gly Pro Val
      35              40              45
gtc tcc ttc tgg ttt ggc agg cgc ctc gtg gtt agt ttg ggc act gtt      297
Val Ser Phe Trp Phe Gly Arg Arg Leu Val Val Ser Leu Gly Thr Val
      50              55              60
gat gta ctg aag cag cat cgg gg      320
Asp Val Leu Lys Gln His Arg
      65              70

```

<210> 200  
 <211> 125  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 40..123

<221> sig\_peptide  
 <222> 40..93  
 <223> Von Heijne matrix  
 score 6.90000009536743  
 seq LELLGSSSPPIISA/SQ

<400> 200  
 cttcctcagt caccaggt ggagtacagt ggcataatc atg gct cac tgc agc 54  
 Met Ala His Cys Ser  
 -15  
 tta gaa ctc ttg ggc tca agc agt cct ccc atc tca gcc tcc caa agc 102  
 Leu Glu Leu Leu Gly Ser Ser Ser Pro Pro Ile Ser Ala Ser Gln Ser  
 -10 -5 1  
 act gga att aca agc gtg agc ca 125  
 Thr Gly Ile Thr Ser Val Ser  
 5 10

<210> 201  
 <211> 210  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 78..209

<221> sig\_peptide  
 <222> 78..128  
 <223> Von Heijne matrix  
 score 6.90000009536743  
 seq LLLLSLSLFLFFW/RQ

<400> 201  
 tcaggttttc ctccttcccg ggtgctctga agtttcacca tgaatcacct tgcaggggct 60  
 ctttttattt tttattg atg ccc agc cag ttg ttg ttg tct ctt tct 110  
 Met Pro Ser Gln Leu Leu Leu Ser Leu Ser  
 -15 -10  
 ctc ttt ttg ttt ttt tgg aga cag agt ctc gtt ttg tgg ccc agg ctg 158  
 Leu Phe Leu Phe Phe Trp Arg Gln Ser Leu Val Leu Trp Pro Arg Leu  
 -5 1 5 10  
 gag tgc agt tgt gtc att gcg gct cac tgc agc ctg acc tcc cag gct 206  
 Glu Cys Ser Cys Val Ile Ala Ala His Cys Ser Leu Thr Ser Gln Ala  
 15 20 25  
 cgg g 210  
 Arg

<210> 202  
 <211> 338  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 89..337

<221> sig\_peptide

<222> 89..226

<223> Von Heijne matrix

score 6.90000009536743

seq CLFCCXFISSCNS/VF

<221> misc\_feature

<222> 291

<223> n=a, g, c or t

<400> 202

```

aattataata atatactaaa atatgtacga atatatacta ataattagta tataatgaat      60
cagtataaaa tatataatat acactaat atg tat act aat aaa tat aca cta      112
                               Met Tyr Thr Asn Lys Tyr Thr Leu
                               -45                               -40
ata tat aac ata cta ata tat aat ata tgt btk drg tat atg tgg ttg      160
Ile Tyr Asn Ile Leu Ile Tyr Asn Ile Cys Xaa Xaa Tyr Met Trp Leu
                               -35                               -25
ata ctc att tat atg tac cta cat att tgc ctc ttt tgt tgc wct ttt      208
Ile Leu Ile Tyr Met Tyr Leu His Ile Cys Leu Phe Cys Cys Xaa Phe
                               -20                               -15                               -10
att tct tcc tgc aat tct gtg ttt ccc tgt gtg att atb ttt ctt ctg      256
Ile Ser Ser Cys Asn Ser Val Phe Pro Cys Val Ile Xaa Phe Leu Leu
                               -5                               1                               5                               10
cct gaa gaa ctt ctt twt gtd twt ctd wdw dtg tnt tty wtt gtg aga      304
Pro Glu Glu Leu Leu Xaa Val Xaa Leu Xaa Xaa Xaa Phe Xaa Val Arg
                               15                               20                               25
tgg agt ctc amt cwg tcg tcc agg ctg gag tgc a      338
Trp Ser Leu Xaa Xaa Ser Ser Arg Leu Glu Cys
                               30                               35

```

<210> 203

<211> 188

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 84..188

<221> sig\_peptide

<222> 84..176

<223> Von Heijne matrix

score 6.90000009536743

seq LWSLIQAVHICLG/RK

<400> 203

```

tattctctga ttckctgtct tggaatgcat ttaaaatctc tgccctgatt ctgacctacc      60
tggcattggga acaagaattt aca atg tta ctc acc cac aat gaa gat tac atg      113
                               Met Leu Leu Thr His Asn Glu Asp Tyr Met
                               -30                               -25
cct ggc aat ttd grc ttw ard daw ttg tgg agc tta att cag gct gtt      161
Pro Gly Asn Xaa Xaa Xaa Xaa Xaa Leu Trp Ser Leu Ile Gln Ala Val
                               -20                               -15                               -10
cat atc tgc cta ggc agg aaa aaa aaa      188
His Ile Cys Leu Gly Arg Lys Lys Lys
                               -5                               1

```

<210> 204

<211> 347

<212> DNA



&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 81..347

&lt;221&gt; sig\_peptide

&lt;222&gt; 81..137

&lt;223&gt; Von Heijne matrix

score 6.90000009536743

seq WVFLVAIIKGVQC/QV

&lt;400&gt; 204

```

agctctggga gaagagcccc agccccagaa ttcccaggag tttccattcg gtgatcagca    60
ctgaacacag aggactcacc atg gag ttt ggg ctg agc tgg gtt ttc ctt gtt    113
                        Met Glu Phe Gly Leu Ser Trp Val Phe Leu Val
                        -15                               -10

gct att ata aaa ggt gtc cag tgt cag gtg caa ctg gtg gag tct ggg    161
Ala Ile Ile Lys Gly Val Gln Cys Gln Val Gln Leu Val Glu Ser Gly
                        -5                               1                               5

ggg ggc ttg gtc aag cct gga ggg tcc cta aga ctc tcc tgt gca gcc    209
Gly Gly Leu Val Lys Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala
                        10                               15                               20

tct gga ttc acc ttc agt gay tac waw atr act kgg att cgc mag gcc    257
Ser Gly Phe Thr Phe Ser Asp Tyr Xaa Xaa Thr Xaa Ile Arg Xaa Ala
                        25                               30                               35                               40

cma ggg aag ggs ytg rak tgg att yca tam atw acg act agt ggg aat    305
Xaa Gly Lys Gly Leu Xaa Trp Ile Xaa Xaa Ile Thr Thr Ser Gly Asn
                        45                               50                               55

acc gca awy tac gca gwc tct gta aag gsc cga ttc acc atc    347
Thr Ala Xaa Tyr Ala Xaa Ser Val Lys Xaa Arg Phe Thr Ile
                        60                               65                               70

```

&lt;210&gt; 205

&lt;211&gt; 440

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 276..440

&lt;221&gt; sig\_peptide

&lt;222&gt; 276..326

&lt;223&gt; Von Heijne matrix

score 6.90000009536743

seq FLFVCLXFDESCS/VT

&lt;400&gt; 205

```

cagtaaat ttttgwwcat tggttccttc dkcttggact ctctgttagc acacctgac    60
agagttggcc gtgttgtaat tctttccctc tctgctgcaa tggtgtttac ttctacctgs    120
caactawkct ttatacttct cttttttgcc atgaagggaac tacatttttt ctttctgtt    180
gggctataca gtgatcctca tcaacaaatt atcaaagaac tgtatgagga aaaggtctct    240
ttttttaaaa gtgaatcagg gctggggagt tagga atg aag agg ttt ttt ttg    293
                        Met Lys Arg Phe Phe Leu
                        -15

ttt gtt tgt ttg tww ttt gac gag tct tgc tct gtc acc agg ctg ggg    341
Phe Val Cys Leu Xaa Phe Asp Glu Ser Cys Ser Val Thr Arg Leu Gly
                        -10                               -5                               1                               5

tgc tgt ggc gcg atc tca gcc cac tgc aam ctc cga ctc cct ggt tca    389
Cys Cys Gly Ala Ile Ser Ala His Cys Xaa Leu Arg Leu Pro Gly Ser
                        10                               15                               20

agc rat dyt cct gcc tca acc tcc cga gta gvy ggg att aca ggc atg    437

```

Ser Xaa Xaa Pro Ala Ser Thr Ser Arg Val Xaa Gly Ile Thr Gly Met  
25 30 35

```
<210> 206
<211> 283
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 162..281

<221> sig_peptide
<222> 162..275
<223> Von Heijne matrix
score 6.90000009536743
seq CMLFVSFLLLLLLG/SR
```

<400>	206		
aataactccc	tttagcattt	cttgtaggac aggtctgatg ttgatgaaat ctctcatctt	60
gtttgtcaga	gaaagtcttt	atcttcctct catgcttgaa ggatgtttcc accggatata	120
ctatcctagg	gtaaaagttt	ttttccttca gcactttaaa t atg tca tgc cac tct	176
		Met Ser Cys His Ser	
		-35	
ctt ctg gcc	tgt aag gtt	ttc act gaa aag tct cct acc aaa cat att	224
Leu Leu Ala	Cys Lys Val Phe Thr	Glu Lys Ser Pro Thr Lys His Ile	
	-30	-25 -20	
aga gag cac	cat tgt atg tta	ttt gtt tct ttt ctc ttg ctg ctt tta	272
Arg Glu His	His Cys Met Leu Phe	Val Ser Phe Leu Leu Leu Leu Leu	
	-15	-10 -5	
gga tcc cg	gg		283
Gly Ser Arg			
1			

```
<210> 207
<211> 264
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 113..262

<221> sig_peptide
<222> 113..190
<223> Von Heijne matrix
      score 6.90000009536743
      seq LLSMLSCCOGACC/PS
```

```

<400> 207
gacggcgggag agcagagagg gagcgcgccct tggctcgctg gccttggcgg cggctctctca    60
ggagagctgg ggcgcccacg agaggatccc tcacccgggt ctctctctcag gg atg aca    118
                                     Met Thr
                                     -25
tca tcc gtc cac ctc ctt gtc ttc aag gac cac ctc ctc tcc atg ctg    166
Ser Ser Val His Leu Leu Val Phe Lys Asp His Leu Leu Ser Met Leu
                                     -20          -15          -10
agc tgc tgc caa ggg gcc tgc tgc cca tct aca cct cac gag ggc act    214
Ser Cys Cys Gln Gly Ala Cys Cys Pro Ser Thr Pro His Glu Gly Thr
                                     -5          1          5
aag aac acg qtt tcc tgg atc cca cca aca tac aaa qca qcc aca caq    262

```

121

Arg Ser Thr Val Ser Trp Ile Pro Pro Thr Tyr Lys Ala Ala Thr Gln  
 10 15 20

gg

264

&lt;210&gt; 208

&lt;211&gt; 422

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 352..420

&lt;221&gt; sig\_peptide

&lt;222&gt; 352..408

&lt;223&gt; Von Heijne matrix

score 6.80000019073486

seq LLSMFVCSHTVQT/AT

&lt;221&gt; misc\_feature

&lt;222&gt; 289..290

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 208

aaaataaaaag tcttcttgat ttccagtgtg ttcctcctgc acwttttggc ctgtttggac 60  
 cacagatttg tggcttttta tgaaatacac ctgtagatta atttwcagtt thtwhayggw 120  
 agtagacagt caaaggctag atcactgtra tgagtagggc ttccacattt aagaaaaagc 180  
 tgtaatgaag tgaattgaat cttgcttctt ttgggtcacc caaaagcagt gataagtgtc 240  
 gagtgtgtta ggcacttatt aacaaaagta actcagaatt gctgtctann cctccatata 300  
 tttttcttc tctccgtgta gttctaaaaa tgaccatatg atattccttg a atg gta 357  
 Met Val

aga gcg tct att ctt ctt agc atg ttc tgt gtg tca cac act gtg cag 405

Arg Ala Ser Ile Leu Leu Ser Met Phe Cys Val Ser His Thr Val Gln  
 -15 -10 -5

aca gca aca tac aca ca

422

Thr Ala Thr Tyr Thr

1

&lt;210&gt; 209

&lt;211&gt; 195

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 39..194

&lt;221&gt; sig\_peptide

&lt;222&gt; 39..89

&lt;223&gt; Von Heijne matrix

score 6.80000019073486

seq ALSSFTWWAPACC/AP

&lt;400&gt; 209

agccactgca cctgggctca cagtttaaata cttgagta atg gag aaa aca gcc ttg 56  
 Met Glu Lys Thr Ala Leu  
 -15

tca tcc ttt acg tgg tgg gca cct gcc tgc tgt gct cca cgt aca tac 104

Ser Ser Phe Thr Trp Trp Ala Pro Ala Cys Cys Ala Pro Arg Thr Tyr  
 -10 -5 1 5

gtg gtg tct gca aca act ctg tca gct gtg caa ggt cac tgt cct cta 152

Val Val Ser Ala Thr Thr Leu Ser Ala Val Gln Gly His Cys Pro Leu

```

                                10                15                        20
cag agt aga aca tcg acc aaa gga aag tta tgg ccg ttt ggg g          195
Gln Ser Arg Thr Ser Thr Lys Gly Lys Leu Trp Pro Phe Gly
      25              30              35

<210> 210
<211> 363
<212> DNA
<213> Homo sapiens


<220>
<221> CDS
<222> 212..361

<221> sig_peptide
<222> 212..280
<223> Von Heijne matrix
       score 6.80000019073486
       seq KLLLSGLTQECLG/AL


<400> 210
taattttcat catctaaact gaatgcaaac agcgcttggc aattaaagt aagctctcca    60
atgaagtata cttcatcagc tgctgtcaag tcatccattg atactgtttt gcgggttttta   120
aattcctttt gtcaactgtga ctgctcatca gcaggcaagg aagagcaggc aaaaaagtt   180
gaaaagtgca tgaaggaaaa ctttgaggaa t atg ata ttc aca ttc cag caa     232
                               Met Ile Phe Thr Phe Gln Gln
                                   -20
att ggg gga aaa ctg cta tta tct ggt tta aca cag gag tgc ctt ggt      280
Ile Gly Gly Lys Leu Leu Leu Ser Gly Leu Thr Gln Glu Cys Leu Gly
        -15             -10             -5
gcc ctg cct gag gct aat gtg ttc tgt agg ggt ggc tgc aca gcc aca     328
Ala Leu Pro Glu Ala Asn Val Phe Cys Arg Gly Gly Cys Thr Ala Thr
1           5           10           15
gtc ctg aaa cat ggg aaa gca tct cct gag tcc ag                      363
Val Leu Lys His Gly Lys Ala Ser Pro Glu Ser
      20              25

<210> 211
<211> 368
<212> DNA
<213> Homo sapiens


<220>
<221> CDS
<222> 230..367

<221> sig_peptide
<222> 230..322
<223> Von Heijne matrix
       score 6.80000019073486
       seq LLALSPLDQAARG/LM


<400> 211
acagagaacc ctgcttcaaa gcagaagtag cagttccgga gtccagctgg ctaaaaactca    60
tccchyggat aatggcaacc catgccttag aaatogettg gctgtttctt ggtgggtgttg   120
gmatgggtgg gsacmrqkwg ggbkgyvack gtcatgcctc agtggdrrag tgtcggcctt   180
cattgaaaac aacatcgtgg tttttgaaaa cttctgggaa ggactgttg atg aat tgc     238
                               Met Asn Cys
                                   -30
gtg agg cag gct aac atc agg atg cag tgc aaa atc tat gat tcc ctg      286
Val Arg Gln Ala Asn Ile Arg Met Gln Cys Lys Ile Tyr Asp Ser Leu
      ..-25             -20             -15
ctg qct ctt tct ccg gac cta cag gca gcc aga ggr ctg atg tgt gct     334

```

123

Leu Ala Leu Ser Pro Asp Leu Gln Ala Ala Arg Gly Leu Met Cys Ala  
 -10 -5 1  
 gct tcc gtg atg tcc ttc ttg gct ttc atg atg g 368  
 Ala Ser Val Met Ser Phe Leu Ala Phe Met Met  
 5 10 15

<210> 212  
 <211> 448  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 316..447

<221> sig\_peptide  
 <222> 316..435  
 <223> Von Heijne matrix  
 score 6.80000019073486  
 seq LLKLLISLRSFWA/ET

<400> 212  
 ttgtcttggc tatacgggat ctttttttggg cccatatgaa atttaagtag cttttcctaa 60  
 ttctgtgaag gaagtcaatg gtagcttgat gggaatagca ttgaatctat aaattacttt 120  
 gggccgtatg gcatttgggc aatattgatt cttcctattc atgagcatgg aatgtttttc 180  
 catttgttca tgctctctct tattttgttg agcagtgggt ttagtctctc cttgaagggg 240  
 ttcttcacat cccttgtaag ttgtattccc aggtatttta ttctctttgt agcaattttg 300  
 aatgggagtt cactc atg att tgg ctc tct ttt tgt cta tta ttg gtg tat 351  
 Met Ile Trp Leu Ser Phe Cys Leu Leu Leu Val Tyr  
 -40 -35 -30  
 agg aat gct tgt gat ttt tgc aca ttg act tta tat cct ggg act ttg 399  
 Arg Asn Ala Cys Asp Phe Cys Thr Leu Thr Leu Tyr Pro Gly Thr Leu  
 -25 -20 -15  
 ctg aag ttg ctt atc agc tta agg agt ttt tgg gct gag acg acg ggg g 448  
 Leu Lys Leu Leu Ile Ser Leu Arg Ser Phe Trp Ala Glu Thr Thr Gly  
 -10 -5 1

<210> 213  
 <211> 158  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 28..156

<221> sig\_peptide  
 <222> 28..102  
 <223> Von Heijne matrix  
 score 6.69999980926514  
 seq LVGSLHLFLSVLA/SK

<400> 213  
 gcgctgggag ttctcttttt cacttga atg ttt tct tct cca ggg ctg agg acg 54  
 Met Phe Ser Ser Pro Gly Leu Arg Thr  
 -25 -20  
 ctc ttt gta ttg gta ggc agc ctg cac ttg ttc ctt tca gtc ctg gca 102  
 Leu Phe Val Leu Val Gly Ser Leu His Leu Phe Leu Ser Val Leu Ala  
 -15 -10 -5  
 agt aaa agc agg aat tct aaa aag caa cga tta ttc ctc cta gtt cct 150  
 Ser Lys Ser Arg Asn Ser Lys Lys Gln Arg Leu Phe Leu Leu Val Pro  
 1 5 10 15  
 ttg tac ag 158

Leu Tyr

<210> 214  
 <211> 193  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 39..191  
 <221> sig\_peptide  
 <222> 39..107  
 <223> Von Heijne matrix  
 score 6.69999980926514  
 seq NVCSLPAPGLCSG/QP

<400> 214  
 aagaaaagct ttgggtcaac tcagcatcat gtttgcag atg ctg aca gac ggg atc 56  
 Met Leu Thr Asp Gly Ile  
 -20  
 cta atg aga gtc aat gtg tgc tca ctg cca gct cct ggg ctg tgc tct 104  
 Leu Met Arg Val Asn Val Cys Ser Leu Pro Ala Pro Gly Leu Cys Ser  
 -15 -10 -5  
 ggt cag cca ggt gtg agg gcc tgg cct ggg gtc aca cag ctg act car 152  
 Gly Gln Pro Gly Val Arg Ala Trp Pro Gly Val Thr Gln Leu Thr Gln  
 1 5 10 15  
 bta gag gaa tgc cca tgg ttc tca gca ttg gaa gga ctg gg 193  
 Xaa Glu Glu Cys Pro Trp Phe Ser Ala Leu Glu Gly Leu  
 20 25

<210> 215  
 <211> 214  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 67..213  
 <221> sig\_peptide  
 <222> 67..165  
 <223> Von Heijne matrix  
 score 6.69999980926514  
 seq ILLLSLIFGPCIL/NS

<400> 215  
 aaagtctcgag aaaatchaga taggcaccaa caagaacgag aaaataacat cccctgggtat 60  
 caaagc atg ttt aac tgg aac cca tgg cta act act tta atc act ggg 108  
 Met Phe Asn Trp Asn Pro Trp Leu Thr Thr Leu Ile Thr Gly  
 -30 -25 -20  
 wta gch gga cct ctc ctc atc cta cta tta agt tta att ttt ggg cct 156  
 Xaa Ala Gly Pro Leu Leu Ile Leu Leu Ser Leu Ile Phe Gly Pro  
 -15 -10 -5  
 tgt ata tta aat tcg ttt ctk aat tkt ata aaa caa cgc ata gct tct 204  
 Cys Ile Leu Asn Ser Phe Leu Asn Xaa Ile Lys Gln Arg Ile Ala Ser  
 1 5 10  
 ggc aaa cgg g 214  
 Gly Lys Arg  
 15

<210> 216  
 <211> 327

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 22..327

&lt;221&gt; sig\_peptide

&lt;222&gt; 22..108

&lt;223&gt; Von Heijne matrix

score 6.69999980926514

seq FCALLLSLXXXXP/XX

&lt;400&gt; 216

```

ctccgcgttc cagaatccaa g atg gcg gga tcc agg caa agg ggt ctc cgg      51
                               Met Ala Gly Ser Arg Gln Arg Gly Leu Arg
                               -25                               -20

gcc aga gtt cgg ccg ctg ttc tgc gcc ttg ctg ctg tca ctm sgw hsv      99
Ala Arg Val Arg Pro Leu Phe Cys Ala Leu Leu Leu Ser Leu Xaa Xaa
                               -15                               -5

mty ckt ccg rkg cka cgs cgt gkg agg aga ccc cgc ggt cgc gtt gcc      147
Xaa Xaa Pro Xaa Xaa Arg Arg Xaa Arg Arg Pro Arg Gly Arg Val Ala
                               1                               5                               10

aca tcg ccg ttt cga gta saa ata cag ctt caa ggg gcc gca cct ggt      195
Thr Ser Pro Phe Arg Val Xaa Ile Gln Leu Gln Gly Ala Ala Pro Gly
                               15                               20                               25

gca gag cga cgg gac cgt gcc ctt ctg ggm cca cgc ggg gaa tgc tat      243
Ala Glu Arg Arg Asp Arg Ala Leu Leu Gly Pro Arg Gly Glu Cys Tyr
                               30                               35                               40                               45

tcc aag ttc aga tca aat tcg agt agc acc atc ttt aaa aag cya aag      291
Ser Lys Phe Arg Ser Asn Ser Ser Ser Thr Ile Phe Lys Lys Xaa Lys
                               50                               55                               60

agg ctc agt gtg gvm aam gac aav agc gga cct ggg      327
Arg Leu Ser Val Xaa Xaa Asp Xaa Ser Gly Pro Gly
                               65                               70

```

&lt;210&gt; 217

&lt;211&gt; 357

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 70..357

&lt;221&gt; sig\_peptide

&lt;222&gt; 70..126

&lt;223&gt; Von Heijne matrix

score 6.69999980926514

seq WVFLVAILKGVHC/DV

&lt;400&gt; 217

```

aggagcccca gccctgggat tcccagctgt ttctgcttgc tgatcaggac tgcacacaga      60
gaactcacc atg gag ttt ggg ctg agc tgg gtt ttc ctt gtt gct att tta      111
                               Met Glu Phe Gly Leu Ser Trp Val Phe Leu Val Ala Ile Leu
                               -15                               -10

aaa ggt gtc cac tgt gac gtg cag ctg gtg gag tcc ggg gga ggt tta      159
Lys Gly Val His Cys Asp Val Gln Leu Val Glu Ser Gly Gly Gly Leu
                               -5                               1                               5                               10

gtt cag cct ggg ggg tcc ctg aga ctc tcc tgt gca gcc tct gga ctc      207
Val Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Leu
                               15                               20                               25

acc ctc agt aac gac tgg atg cac tgg gtc cgc caa gcc cca ggg aag      255

```

126

Thr	Leu	Ser	Asn	Asp	Trp	Met	His	Trp	Val	Arg	Gln	Ala	Pro	Gly	Lys	
	30					35					40					
ggg	ctg	gtg	tgg	gtc	tca	cac	att	gat	agt	tct	rgg	act	atc	aca	aat	303
Gly	Leu	Val	Trp	Val	Ser	His	Ile	Asp	Ser	Ser	Xaa	Thr	Ile	Thr	Asn	
	45					50					55					
tac	gcg	gac	tcc	gtg	aag	ggc	cga	ttc	acc	atc	tcc	aga	gac	aac	gcc	351
Tyr	Ala	Asp	Ser	Val	Lys	Gly	Arg	Phe	Thr	Ile	Ser	Arg	Asp	Asn	Ala	
60					65					70					75	
aag	tgg															357
Lys	Trp															

&lt;210&gt; 218

&lt;211&gt; 189

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 74..187

&lt;221&gt; sig\_peptide

&lt;222&gt; 74..118

&lt;223&gt; Von Heijne matrix

score 6.69999980926514

seq LFLGFLACSVAYQ/CH

&lt;400&gt; 218

ttatcaagga	cactgtcttt	tcgccatcat	gtgttcttgg	cccctctgtt	gaaattcaat	60
ctatcataga	caa atg ggt tta ttt ctg ggc ttt cta gcc tgt tct gtt	109				
	Met Gly Leu Phe Leu Gly Phe Leu Ala Cys Ser Val					
	-15	-10	-5			
gca tac cag tgc cat tct gct ttt gtt act gta gct tca cag tac act	157					
Ala Tyr Gln Cys His Ser Ala Phe Val Thr Val Ala Ser Gln Tyr Thr						
	1	5	10			
ttg aaa tca gag act ttg atg ccc gca gcg gg	189					
Leu Lys Ser Glu Thr Leu Met Pro Ala Ala						
15	20					

&lt;210&gt; 219

&lt;211&gt; 353

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 41..352

&lt;221&gt; sig\_peptide

&lt;222&gt; 41..187

&lt;223&gt; Von Heijne matrix

score 6.69999980926514

seq FLGLIFFLELATG/IL

&lt;400&gt; 219

agtgacttg	ccatctgcct	tgcaggatgg	catccagccc	atg tgg gag gac agc	55
				Met Trp Glu Asp Ser	
				-45	
agg aat aaa cgg ggt ggc cgc tgg ctg gtc agc ctg gcc aag cag cag	103				
Arg Asn Lys Arg Gly Gly Arg Trp Leu Val Ser Leu Ala Lys Gln Gln					
	-40	-35	-30		
cgc cac att gag ctg gac cgg ctg tgg ctg gag acg ttc tcc gtg ttc	151				
Arg His Ile Glu Asp Arg Leu Trp Leu Glu Thr Phe Ser Val Phe					
-25	-20	-15			



127

```

ctc ggt ctc atc ttc ttc ctg gag ctg gca aca ggg atc ctg gcc ttt      199
Leu Gly Leu Ile Phe Phe Leu Glu Leu Ala Thr Gly Ile Leu Ala Phe
      -10                -5                1
gtc ttc aag gac tgg att cga gac cag ctc aac ctc ttc atc aac aac      247
Val Phe Lys Asp Trp Ile Arg Asp Gln Leu Asn Leu Phe Ile Asn Asn
      5                10                15                20
aac gtc aag gcc tac cgg gac gac att gac ctc cag arc ctc att gac      295
Asn Val Lys Ala Tyr Arg Asp Asp Ile Asp Leu Gln Xaa Leu Ile Asp
      25                30                35
ttt gct cag gaa tac tgg tct tgc tgc gga scc gag gcc cca ata rdt      343
Phe Ala Gln Glu Tyr Trp Ser Cys Cys Gly Xaa Glu Ala Pro Ile Xaa
      40                45                50
gga acc ggg g
Gly Thr Gly
      55

```

<210> 220  
 <211> 115  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 12..113

<221> sig\_peptide  
 <222> 12..53  
 <223> Von Heijne matrix  
 score 6.59999990463257  
 seq FLSLSTAFWVYA/MI

```

<400> 220
actagcatTTT c atg ttt tta tct ctc tct act gca ttc tgg gta gtt tat      50
      Met Phe Leu Ser Leu Ser Thr Ala Phe Trp Val Val Tyr
      -10                -5
gcc atg ata att tat tca gct ctc tct gct gga ttt att att ttc ttt      98
Ala Met Ile Ile Tyr Ser Ala Leu Ser Ala Gly Phe Ile Ile Phe Phe
      1                5                10                15
tta gtt gtg ttt aat ct
Leu Val Val Phe Asn
      20

```

<210> 221  
 <211> 142  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 29..142

<221> sig\_peptide  
 <222> 29..130  
 <223> Von Heijne matrix  
 score 6.59999990463257  
 seq FFLFCFETGSHS/VT

```

<400> 221
cctgccccatt gcttcaacct gcacctct atg tac att gtg atg gat cta cct      52
      Met Tyr Ile Val Met Asp Leu Pro
      -30
cta tgg ctc tcc cat gag gtc caa tct tat att cct tct ttc ttc ctt      100
Leu Trp Leu Ser His Glu Val Gln Ser Tyr Ile Pro Ser Phe Phe Leu

```

128

```

      -25              -20              -15
    ttt ttt tgc ttt gag act ggg tct cac tct gtc acc cac ggg      142
    Phe Phe Cys Phe Glu Thr Gly Ser His Ser Val Thr His Gly
    -10              -5              1

```

<210> 222  
 <211> 370  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 209..370

<221> sig\_peptide  
 <222> 209..289  
 <223> Von Heijne matrix  
       score 6.59999990463257  
       seq LAFSFSFFPSSFS/SF

```

<400> 222
ttttggatcac atagactca tgtgacatta gatcacagca tttttgtttt tattattaat      60
atattgcctt agaactacat tgctaaacct ggtcttttgta tctgcgaagt tctaaccatct    120
tgccacagct tagttagctt tgagagggaa agggtagaat ccatttaagg agacaggtta      180
aaaaatgata tatttaagca tataggca atg gta gca cat gat tac caa aac          232
                               Met Val Ala His Asp Tyr Gln Asn
                               -25              -20
ata att agc ctt ttc ttt ctt gct ttt tca ttt tct ttc ttt cct tct      280
Ile Ile Ser Leu Phe Phe Leu Ala Phe Ser Phe Ser Phe Phe Pro Ser
                               -15              -10              -5
tca ttt tct tct ttc ttt ctt ktc ttt ctt tct ttt ttc tct tct ttc      328
Ser Phe Ser Ser Phe Phe Leu Xaa Phe Leu Ser Phe Phe Ser Ser Phe
                               1              5              10
ttt ctc tct ctt ctt tct ttc cct tcc ttc ctc ccc ccc ggr              370
Phe Leu Ser Leu Leu Ser Phe Pro Ser Phe Leu Pro Pro Gly
                               15              20              25

```

<210> 223  
 <211> 431  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 22..429

<221> sig\_peptide  
 <222> 22..66  
 <223> Von Heijne matrix  
       score 6.59999990463257  
       seq ALRALCGFRGVAA/QV

```

<400> 223
gcagtctgca gccggagtaa g atg gcg gcg ctg agg gct ttg tgc ggc ttc      51
                               Met Ala Ala Leu Arg Ala Leu Cys Gly Phe
                               -15              -10
cgg ggc gtc gcg gcc cag gtg ctg cgg mct ggg gct gga gtc cga ttg      99
Arg Gly Val Ala Ala Gln Val Leu Arg Xaa Gly Ala Gly Val Arg Leu
-5              1              5              10
ccg att cag ccc agc aga ggt gtt cgg cag tgg cag cca gat gtg gaa      147
Pro Ile Gln Pro Ser Arg Gly Val Arg Gln Trp Gln Pro Asp Val Glu
                               15              20              25
tgg gca cag cag ttt ggg gga gct gtt atg tac cca agc aaa gaa aca      195

```

Trp	Ala	Gln	Gln	Phe	Gly	Gly	Ala	Val	Met	Tyr	Pro	Ser	Lys	Glu	Thr	
	30						35					40				
gcc	cac	tgg	aag	cct	cca	cct	tgg	aat	gat	gtg	gac	cct	cca	aag	gac	243
Ala	His	Trp	Lys	Pro	Pro	Pro	Trp	Asn	Asp	Val	Asp	Pro	Pro	Lys	Asp	
	45						50					55				
aca	att	gtg	aag	aac	att	acc	ctg	aac	ttt	ggg	ccc	caa	cac	cca	gca	291
Thr	Ile	Val	Lys	Asn	Ile	Thr	Leu	Asn	Phe	Gly	Pro	Gln	His	Pro	Ala	
	60						65			70					75	
gcg	cat	ggt	gtc	ctg	cga	cta	gtg	atg	gaa	ttg	agt	ggg	gag	atg	gtg	339
Ala	His	Gly	Val	Leu	Arg	Leu	Val	Met	Glu	Leu	Ser	Gly	Glu	Met	Val	
				80					85					90		
cgg	aag	tgt	gat	cct	cac	atc	ggg	ctc	ctg	cac	cga	ggc	act	gag	aag	387
Arg	Lys	Cys	Asp	Pro	His	Ile	Gly	Leu	Leu	His	Arg	Gly	Thr	Glu	Lys	
			95					100					105			
ctc	att	gaa	tac	aag	rcc	tat	ctt	cag	gcc	ctt	cca	tac	ttt	ga		431
Leu	Ile	Glu	Tyr	Lys	Xaa	Tyr	Leu	Gln	Ala	Leu	Pro	Tyr	Phe			
		110						115				120				

```
<210> 224
<211> 282
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 132..281
```

```
<221> sig_peptide
<222> 132..215
<223> Von Heijne matrix
      score 6.5
      seq LVFLHLFLXVYLG/LV
```

```

<400> 224
atttttaaagt gtttctgtta atgtattcta cttcagtcgcc ccaaaattcc aactaacgac      60
atacatgaat aacagatcat gactgctggt tctacaagcc tttctgctca ctgtgcttcc      120
acttacaact c atg tta ata tgg tct tcc tct tct ttt cct gca ccc cct      170
               Met Leu Ile Trp Ser Ser Ser Ser Phe Pro Ala Pro Pro
                   -25                               -20

ctc ttt ctt gtc ttt ctt cat ctt ttc ctt mwt gtc tat ttg ggt ctt      218
Leu Phe Leu Val Phe Leu His Leu Phe Leu Xaa Val Tyr Leu Gly Leu
-15                               -10                               -5                               1

gtc atg ccc act caa cag tat ctc ctc ctg cag agt cca ttg atg ttc      266
Val Met Pro Thr Gln Gln Tyr Leu Leu Leu Gln Ser Pro Leu Met Phe
               5                               10                               15

aca gac aaa gcc cag c      282
Thr Asp Lys Ala Gln
               20

```

```
<210> 225
<211> 198
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 26..196
```

```
<221> sig_peptide
<222> 26..163
<223> Von Heijne matrix
      score 6.5
      seq WLFFLMSLCTPP/DR
```



131

Gly	Lys	Arg	Gly	Gln	Ala	Trp	Arg	Leu	Met	Pro	Val	Val	Pro	Ala	Val		
1				5					10						15		
tgg	gag	cct	gag	gca	ggg	gga	ttg	ctt	cag	ctc	ggg	ggg	tct	agg	g		206
Trp	Glu	Pro	Glu	Ala	Gly	Gly	Leu	Leu	Gln	Leu	Gly	Gly	Ser	Arg			
				20					25					30			

<210> 228  
 <211> 480  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 216..479

<221> sig\_peptide  
 <222> 216..326  
 <223> Von Heijne matrix  
 score 6.5  
 seq LLVFFLIVRTLSC/RS

<400> 228																	
gcatccccck	ktagctcaga	gaagtttggt	rdgaccgatc	ttctgaagcc	tacttctgtc												60
aactcatcaa	agtcattctc	catccagctt	tgttccatta	tgggtgagga	gctacgatcc												120
tttgaggag	aagaggcact	ctgattttta	gaattttcag	cttttctgct	ctgggttcgc												180
cccatctttg	tgggtttatc	taccttcggt	ctttg atg	atg gtg	acc tac	aga											233
			Met Met	Val Thr	Tyr Arg												
			-35														
tgg ggt ttt	ggt gtg	gat gtc	mtt ttt	ggt gct	ggt gat	gct att	cct										281
Trp Gly Phe	Gly Val	Asp Val	Xaa Phe	Val Ala	Val Asp	Ala Ile	Pro										
-30		-25		-20													
ttc tgt ttg	tta gtt	ttc ttt	cta ata	gtc agg	acc ctc	agc tgc	agg										329
Phe Cys Leu	Leu Val	Phe Phe	Leu Ile	Val Arg	Thr Leu	Ser Cys	Arg										
-15		-10		-5		1											
tct gtt gga	gta tgc	tgg agg	tcc act	cca gac	cct gtt	tgc cta	ggg										377
Ser Val Gly	Val Cys	Trp Arg	Ser Thr	Pro Asp	Pro Val	Cys Leu	Gly										
	5		10		15												
atc acc agc	aga ggc	tgc aga	aca gaa	ata ttg	cag aac	agc aaa	tgt										425
Ile Thr Ser	Arg Gly	Cys Arg	Thr Glu	Ile Leu	Gln Asn	Ser Lys	Cys										
	20		25		30												
tgc tcc ctg	atc ctt	cct ctg	gaa gct	tcg tct	caa agg	ggc act	gaa										473
Cys Ser Leu	Ile Leu	Pro Leu	Glu Ala	Ser Ser	Gln Arg	Gly Thr	Glu										
	35		40		45												
tgt atg a																	480
Cys Met																	
50																	

<210> 229  
 <211> 144  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 43..144

<221> sig\_peptide  
 <222> 43..99  
 <223> Von Heijne matrix  
 score 6.5  
 seq EIFLPFSLSPANA/QS

<400> 229

132

```

tccagatgtg atttgggtatt tcatactttg ttgcttttgt aa atg ctg tac cca      54
                                Met Leu Tyr Pro
ctg cct gag ata ttc tta cct ttt tct ttg tcc cca gca aat gcc cag      102
Leu Pro Glu Ile Phe Leu Pro Phe Ser Leu Ser Pro Ala Asn Ala Gln
-15                -10                -5                1
tca aaa ttt agc ctt tat ttt ttt ccc ttg gtg aag ccg ggg      144
Ser Lys Phe Ser Leu Tyr Phe Phe Pro Leu Val Lys Pro Gly
          5                10                15

```

&lt;210&gt; 230

&lt;211&gt; 457

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 314..457

&lt;221&gt; sig\_peptide

&lt;222&gt; 314..394

&lt;223&gt; Von Heijne matrix

score 6.40000009536743

seq RLLCLXFXRLLLG/TS

&lt;221&gt; misc\_feature

&lt;222&gt; 118,258..259,303,440

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 230

```

agctccgcgg taatggaggg tagggatggg tgctgaagta tcaggctctg gctctagctt      60
tagctctggc actggaactg cgtcggagtc tgggtctgag tctggcagcc cgaagcctg      120
grmcaccttt tcttgattct ctaaggcggg ggctgcctgc gtccaagcag ctggtttgca      180
gcgttccaac gctgggaggg agttccctta cctgggggtcc agtctgtaaa gttgtcgccg      240
ctttctaggg acccgccnnd scggctggga ctcttccatg cgtgagtatt actgarstgc      300
tsnaaggtcc ggc atg tcc ctg gaa cct gcc tcg gsc ctc ttg ggt gtg      349
          Met Ser Leu Glu Pro Ala Ser Xaa Leu Leu Gly Val
          -25                -20
cgg cgg aga ctg ctt tgt cta mct ttc tsc cga ctt ctc tta ggr acc      397
Arg Arg Arg Leu Leu Cys Leu Xaa Phe Xaa Arg Leu Leu Leu Gly Thr
-15                -10                -5                1
agt ctg ttg aag ttt gtg gkc tcc tgs agy cca ccc ama ccg nat act      445
Ser Leu Leu Lys Phe Val Xaa Ser Xaa Ser Pro Pro Xaa Pro Xaa Thr
          5                10                15
ctc acc tct tcc      457
Leu Thr Ser Ser
          20

```

&lt;210&gt; 231

&lt;211&gt; 112

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 12..110

&lt;221&gt; sig\_peptide

&lt;222&gt; 12..83

&lt;223&gt; Von Heijne matrix

score 6.40000009536743

seq LSVLILCVCVCVC/VY

<400> 231  
 ctgattttkc t atg ytg att ttg tat ctk gca act tta cta aat tta tca 50  
                   Met Leu Ile Leu Tyr Leu Ala Thr Leu Leu Asn Leu Ser  
                                   -20                                  -15  
 gtt cta ata ctt tgt gtg tgt gtg tgt gtg tgt gtg tat gat tta tat 98  
 Val Leu Ile Leu Cys Val Cys Val Cys Val Cys Val Tyr Asp Leu Tyr  
                   -10                  -5                                  1                                  5  
 ata waa agg gga gt 112  
 Ile Xaa Arg Gly

<210> 232  
 <211> 359  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 8..358  
 <221> sig\_peptide  
 <222> 8..55  
 <223> Von Heijne matrix  
           score 6.40000009536743  
           seq LGTTVLLWSLLRS/SP

<221> misc\_feature  
 <222> 326  
 <223> n=a, g, c or t

<400> 232  
 gataatc atg gcg ccc ctc gga aca act gta ttg ctg tgg agc ctc ttg 49  
           Met Ala Pro Leu Gly Thr Thr Val Leu Leu Trp Ser Leu Leu  
                   -15                                  -10                                  -5  
 agg agt tct ccg ggc gtg gaa cgg gtc tgt ttc cgg gct cga atc cag 97  
 Arg Ser Ser Pro Gly Val Glu Arg Val Cys Phe Arg Ala Arg Ile Gln  
           1                                  5                                  10  
 ccc tgg cac ggt ggc ctg ctc caa ccg cta cct tgc tct ttc gag atg 145  
 Pro Trp His Gly Gly Leu Leu Gln Pro Leu Pro Cys Ser Phe Glu Met  
           15                                  20                                  25                                  30  
 ggg ctg cca cgc cgc cgg ttc agc tcc gag gcc gca gaa tct ggt agc 193  
 Gly Leu Pro Arg Arg Arg Phe Ser Ser Glu Ala Ala Glu Ser Gly Ser  
                                   35                                  40                                  45  
 .cca gag acc aag aaa cct aca ttt atg gat gag gaa gtt caa agc ata 241  
 Pro Glu Thr Lys Lys Pro Thr Phe Met Asp Glu Glu Val Gln Ser Ile  
                                   50                                  55                                  60  
 ctc acg aaa atg aca ggc ttg aac ttg cag aag act ttt aag cca gct 289  
 Leu Thr Lys Met Thr Gly Leu Asn Leu Gln Lys Thr Phe Lys Pro Ala  
                   65                                  70                                  75  
 ata caa gaa ctg aag cca cca acc tat aag cta atg nct cag gca cag 337  
 Ile Gln Glu Leu Lys Pro Pro Thr Tyr Lys Leu Met Xaa Gln Ala Gln  
           80                                  85                                  90  
 ttg gaa gag gct aca aga cag g 359  
 Leu Glu Glu Ala Thr Arg Gln  
           95                                  100

<210> 233  
 <211> 301  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS

&lt;222&gt; 4..300

&lt;221&gt; sig\_peptide

&lt;222&gt; 4..105

&lt;223&gt; Von Heijne matrix

score 6.40000009536743

seq LLFVVLLPPPPGS/VX

&lt;221&gt; misc\_feature

&lt;222&gt; 124,129,162

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 233

gcg atg ctc ctc act ttc agc tcc agc tcc cgc cac cgc cgc ctc tat	48
Met Leu Leu Thr Phe Ser Ser Ser Ser Arg His Arg Arg Leu Tyr	
-30 -25 -20	
cgc cgc cgc cgc cac cac ctc ctc ttc gtt gtc ctc ctt cct cct ccg	96
Arg Arg Arg Arg His His Leu Leu Phe Val Val Leu Leu Pro Pro Pro	
-15 -10 -5	
cct ggc agc gtt gkt ctc tgc agc sgg nrm grn smv raa gtg ctr vbg	144
Pro Gly Ser Val Xaa Leu Cys Ser Xaa Xaa Xaa Xaa Xaa Val Leu Xaa	
1 5 10	
kma sga aag ttc cgg gan gga cta cat gga gcc atg ctc cct ggg ctc	192
Xaa Xaa Lys Phe Arg Xaa Gly Leu His Gly Ala Met Leu Pro Gly Leu	
15 20 25	
ttc cgc ggg cgc ccg cgc gct gcc ctt cgc ttg aga gtc tca ccg wgt	240
Phe Arg Gly Arg Pro Arg Ala Ala Leu Arg Leu Arg Val Ser Pro Xaa	
30 35 40 45	
tgc cca ggc tgg aaa gtg gcg cga tct cgg ctc aca gca acc tcc gcc	288
Cys Pro Gly Trp Lys Val Ala Arg Ser Arg Leu Thr Ala Thr Ser Ala	
50 55 60	
tcm cgg gmc cgg g	301
Ser Arg Xaa Arg	
65	

&lt;210&gt; 234

&lt;211&gt; 248

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 152..247

&lt;221&gt; sig\_peptide

&lt;222&gt; 152..190

&lt;223&gt; Von Heijne matrix

score 6.40000009536743

seq MLLLLQLNLKTLS/SS

&lt;400&gt; 234

acaagttggg tgctgtcgcc tgcgcacgcg cgggcccggga ggctgagcag tactgttgag	60
agcgggtgtga ggtgcttggt agcgcgccgt agctgcttcc acgtccttgc ttcacctcag	120
gtaaagagag aagtaatgga aggcctgtct g atg ttg ctt ctt ttg caa cta	172
Met Leu Leu Leu Leu Gln Leu	
-10	
aac tta aaa aca ctc tca tcc agt acc ata gca ttg aag aag ata agt	220
Asn Leu Lys Thr Leu Ser Ser Ser Thr Ile Ala Leu Lys Lys Ile Ser	
-5 1 5 10	
ggc gag ttg cta aga aaa cga aag agg g	248
Gly Glu Leu Leu Arg Lys Arg Lys Arg	
15	



<210> 235  
 <211> 393  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 339..392

<221> sig\_peptide  
 <222> 339..383  
 <223> Von Heijne matrix  
 score 6.40000009536743  
 seq LFVLLIITQLLYG/GI

<400> 235  
 gttccaaagt gagctgtctc tggcagcatt catatagaat agaatttgaa tgggtgcaccc 60  
 agatttgaac aacatggtaa tcatgtgatg gacatggaaa agtgractaa cbtkrgggat 120  
 cwtgggtargg tcaytaagaa taactckaat cawgatgtta aaaggctttc ctttacattc 180  
 acaaaaacaat ttrsttccta gaagtagttt attccttgct gtggtcattt ttgctccttt 240  
 ataatactac atctaaatca atttgttaaa tatagtagag aaatgaaata aatttcttcc 300  
 agttaaacca ctgcacttaa agagtagaaa ccctctct atg tca ctc ttt gtt ttg 356  
 Met Ser Leu Phe Val Leu  
 -15 -10  
 ttg atc ata act caa ctg ctg tat ggt ggg ata ctc t 393  
 Leu Ile Ile Thr Gln Leu Leu Tyr Gly Gly Ile Leu  
 -5 1

<210> 236  
 <211> 222  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 121..222

<221> sig\_peptide  
 <222> 121..204  
 <223> Von Heijne matrix  
 score 6.40000009536743  
 seq ILFLGVLLSASDL/CV

<400> 236  
 ttttgagtta atttttgtat aagttgtaag gattaggtca gggttcttaa gaaaaatatt 60  
 gttttggtct atagatgtct cattgcttct gtgctatttg ttggaaaagc tgttcttcca 120  
 atg aat tgc ttt tgc aat ttt gtc aaa acc agt gag gca tat atg att 168  
 Met Asn Cys Phe Cys Asn Phe Val Lys Thr Ser Glu Ala Tyr Met Ile  
 -25 -20 -15  
 ctg ttt cta ggt gtt cta ctc tct gca agt gat tta tgt gtc tat ccc 216  
 Leu Phe Leu Gly Val Leu Leu Ser Ala Ser Asp Leu Cys Val Tyr Pro  
 -10 -5 1  
 atc ggg 222  
 Ile Gly  
 5

<210> 237  
 <211> 154  
 <212> DNA  
 <213> Homo sapiens

<220>

136

<221> CDS  
 <222> 54..152

<221> sig\_peptide  
 <222> 54..95  
 <223> Von Heijne matrix  
 score 6.40000009536743  
 seq SVILALWEAEAGG/SP

<400> 237  
 agtcttttgc tcctgtgggt aagattattc tgctaggctg ctcacgggtg ctg atg 56  
 Met  
 tct gta atc cta gca ctt tgg gag gcc gag gcg ggc gga tcg cct gag 104  
 Ser Val Ile Leu Ala Leu Trp Glu Ala Glu Ala Gly Gly Ser Pro Glu  
 -10 -5 1  
 atc ggg agt tcg gga ccg gcc gca cca aca tgg aga agc ccc gtc cag 152  
 Ile Gly Ser Ser Gly Pro Ala Ala Pro Thr Trp Arg Ser Pro Val Gln  
 5 10 15  
 gg 154

<210> 238  
 <211> 439  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 255..437  
 <221> sig\_peptide  
 <222> 255..341  
 <223> Von Heijne matrix  
 score 6.30000019073486  
 seq LGCLLLAVRSSAT/VN

<221> misc\_feature  
 <222> 359..360,381  
 <223> n=a, g, c or t

<400> 238  
 tcaccacaat caatttttaga acattttcat catcccgaaa ataagccctg ttccctttag 60  
 ctgtcactcc ccactcctac cccccagccc tgtgcaataa tctactttct gtctttgaag 120  
 ctttgccctat tctggacatt ttgtataaaa gggtttgtgg aggatgtggt cttttgtgac 180  
 tggtcttcttg aacttggcat agtgttttca aggttcaacc atgttgtagc acgtacgttc 240  
 ctttttatgg ccaa atg tac gga gag tcc aca ttg ttt atc cat tca tca 290  
 Met Tyr Gly Glu Ser Thr Leu Phe Ile His Ser Ser  
 -25 -20  
 gtt cat ggg cat ttg ggt tgt ctc ctc ttg gct gtt agg agt agt gct 338  
 Val His Gly His Leu Gly Cys Leu Leu Leu Ala Val Arg Ser Ser Ala  
 -15 -10 -5  
 act gtg aac att acg tac chn nkw gtk tgt gtg gac att cak ntt cat 386  
 Thr Val Asn Ile Thr Tyr Xaa Xaa Val Cys Val Asp Ile Xaa Xaa His  
 1 5 10 15  
 ttc cat atg ctt atg tct gga att act ggg tca tat ggc aac tct ctt 434  
 Phe His Met Leu Met Ser Gly Ile Thr Gly Ser Tyr Gly Asn Ser Leu  
 20 25 30  
 tca ct 439  
 Ser

<210> 239  
 <211> 229  
 <212> DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 7..228

&lt;221&gt; sig\_peptide

&lt;222&gt; 7..159

&lt;223&gt; Von Heijne matrix

score 6.30000019073486

seq WLYLLEVVAPLSG/IH

&lt;400&gt; 239

```

gtcaag atg gcg gcg tct gta tta aac acc gtg ctg agg cgg ctt cct      48
      Met Ala Ala Ser Val Leu Asn Thr Val Leu Arg Arg Leu Pro
          -50                      -45                      -40
atg cta tct ctc ttc cga ggt tct cac aga gtt cag gta act ctt cga      96
Met Leu Ser Leu Phe Arg Gly Ser His Arg Val Gln Val Thr Leu Arg
      -35                      -30                      -25
aag aca ttt tgc aca acc tca agt tgg tta tac ctt ctc gag gtt gtc     144
Lys Thr Phe Cys Thr Thr Ser Ser Trp Leu Tyr Leu Leu Glu Val Val
      -20                      -15                      -10
gct cca ctg tca gga atc cac gag tgg aga cct tcc cac gtg tgt ctt     192
Ala Pro Leu Ser Gly Ile His Glu Trp Arg Pro Ser His Val Cys Leu
      -5                      1                      5                      10
agc tgt cta ggc agt act tcc tgc aac ccc cct gag g                      229
Ser Cys Leu Gly Ser Thr Ser Cys Asn Pro Pro Glu
          15                      20

```

&lt;210&gt; 240

&lt;211&gt; 318

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 65..316

&lt;221&gt; sig\_peptide

&lt;222&gt; 65..259

&lt;223&gt; Von Heijne matrix

score 6.30000019073486

seq LMVVAETSQGSWS/AP

&lt;221&gt; misc\_feature

&lt;222&gt; 259

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 240

```

ctcttcgggt gtccagccct tctcccagcc ctggtccctc agaaggaggg taactccctt      60
ccag atg tta cgg tcc gcc tgc gtc tct cag cac gcc ggt ggc att tgg     109
      Met Leu Arg Ser Ala Cys Val Ser Gln His Ala Gly Gly Ile Trp
          -65                      -60                      -55
gtt gac cgc gga ggc ccc cag tgc cag agg gtg ttc acg ttc tgc cgt     157
Val Asp Arg Gly Gly Pro Gln Cys Gln Arg Val Phe Thr Phe Cys Arg
      -50                      -45                      -40                      -35
ggg ctc agc cca aac ttt gga cgc tca gag acc caa cgg gag cgc tgg     205
Gly Leu Ser Pro Asn Phe Gly Arg Ser Glu Thr Gln Arg Glu Arg Trp
          -30                      -25                      -20
ata agg cca gga cag ctg atg gtt gtg gca gaa aca tct caa ggt agc     253
Ile Arg Pro Gly Gln Leu Met Val Val Ala Glu Thr Ser Gln Gly Ser
          -15                      -10                      -5

```

138

tgg tcn gcc ccc act tcc cca tst acc tct tgt cct ccc ccc aac acc 301  
 Trp Ser Ala Pro Thr Ser Pro Xaa Thr Ser Cys Pro Pro Pro Asn Thr  
           1                  5                  10  
 asc acc aca ccg gyt cc 318  
 Xaa Thr Thr Pro Xaa  
 15

<210> 241  
 <211> 405  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 123..404

<221> sig\_peptide  
 <222> 123..257  
 <223> Von Heijne matrix  
       score 6.30000019073486  
       seq GFVSLLVHAADA/WV

<400> 241  
 tagctggacc cgtctgggag gtaggtttgt gagcgtgaga gaks gatctg taccgcgggg 60  
 atccgaagta tgcttatcca ggtgggctgc ctcaagcctc gatcccaccc ccgcgctdvt 120  
 ag atg gtg tca agg tcc ttg cgt ggg aga agg act tgg gtg aga tgc 167  
   Met Val Ser Arg Ser Leu Arg Gly Arg Arg Thr Trp Val Arg Cys  
   -45                  -40                  -35  
 atg cgg aga ttg ccc cca att ccg gcc tgg agc caa ggg aaa ggg atg 215  
 Met Arg Arg Leu Pro Pro Ile Pro Ala Trp Ser Gln Gly Lys Gly Met  
 -30                  -25                  -20                  -15  
 cct gga ttt gtg tct cta ttg gtg gtc cac gct gcg gat gcc tgg gta 263  
 Pro Gly Phe Val Ser Leu Leu Val Val His Ala Ala Asp Ala Trp Val  
                   -10                  -5                  1  
 gcc cag agg ttr tct acg cca tac ttc tca ctg ttt ttg agc ata cct 311  
 Ala Gln Arg Leu Ser Thr Pro Tyr Phe Ser Leu Phe Leu Ser Ile Pro  
   5                  10                  15  
 aga tgt tcc ttt cct agg cgg agt ata gat cgc acg tgt tct agc stc 359  
 Arg Cys Ser Phe Pro Arg Arg Ser Ile Asp Arg Thr Cys Ser Ser Xaa  
   20                  25                  30  
 tta gac tca gag ggt tcg agc tct ata asc ccc tcc act ccc ttc a 405  
 Leu Asp Ser Glu Gly Ser Ser Ser Ile Xaa Pro Ser Thr Pro Phe  
 35                  40                  45

<210> 242  
 <211> 242  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 129..242

<221> sig\_peptide  
 <222> 129..191  
 <223> Von Heijne matrix  
       score 6.30000019073486  
       seq SLLPCSLISDCCA/SN

<400> 242  
 cttttgtttt gcaatgccct gccccagag gtggagteta cagaggcagg caggcctcct 60  
 tgagctgagg tgggtccac ccagttcgag cttccagct gctttgttta cctactcaag 120  
 cctgggca atg gtg ggc gcc ctt ccc cca gcc tcg ctt ctg cct tgc agt 170

139

Met Val Gly Ala Leu Pro Pro Ala Ser Leu Leu Pro Cys Ser  
 -20 -15 -10  
 ttg atc tca gac tgc tgt gct agc aat gag cga ggc tcc atg ggc gta 218  
 Leu Ile Ser Asp Cys Cys Ala Ser Asn Glu Arg Gly Ser Met Gly Val  
 -5 1 5  
 gga ccc tct gag cca cgg cgy ggg 242  
 Gly Pro Ser Glu Pro Arg Arg Gly  
 10 15

<210> 243  
 <211> 363  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 298..363

<221> sig\_peptide  
 <222> 298..357  
 <223> Von Heijne matrix  
 score 6.30000019073486  
 seq LGSLLASLAPSTG/LG

<400> 243  
 accactctga ggagacgcgt gacagataag aagggtcgtt gggatcagtc ctggtggttag 60  
 cttaggaagc agagcctgga gcatctccac tatggcctgg gctccactac ttctcaccct 120  
 cctcgctcac tgcacaggtt cttgggcca ctttatgctg actcagccgc actctgtgtc 180  
 ggagtcgccc gssgaagacg gtaaccatct cctgcacccg cagcagtggtc agctttgtca 240  
 gcaactatgt tcagtgggtac cagcggcgcc cggacagtgc cccaccact gtgatct 297  
 atg agg atg aca aaa gac cct ctg ggg tct ctg atc gct tct ctg gct 345  
 Met Arg Met Thr Lys Asp Pro Leu Gly Ser Leu Ile Ala Ser Leu Ala  
 -20 -15 -10 -5  
 cca tcg aca ggt ctt ggg 363  
 Pro Ser Thr Gly Leu Gly  
 1

<210> 244  
 <211> 324  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 153..323

<221> sig\_peptide  
 <222> 153..236  
 <223> Von Heijne matrix  
 score 6.30000019073486  
 seq FFLFLFFXEXXX/XX

<400> 244  
 aattgatact gctttagatg tttctgtctc attttacaaa aatgtaagaa aaaagaaaaa 60  
 tcaaactata ctgttaccta tttctgttat attcttaaca gaatgttctg tacacataag 120  
 tgtatgtgtg ttaatcctct tgtttaaagt cc atg aaa ctt cag ttt gcc ttt 173  
 Met Lys Leu Gln Phe Ala Phe  
 -25  
 tgt tat ttt ctt tat tta gat acc ttt ttt ctt ttt ttt ttt ttt ttt 221  
 Cys Tyr Phe Leu Tyr Leu Asp Thr Phe Phe Leu Phe Leu Phe Phe Xaa  
 -20 -15 -10  
 gag ama gyc tkg cyc kgt kgc hta ggm agg agt gca gtg gca maa cct 269  
 Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly Arg Ser Ala Val Ala Xaa Pro

2.

141

Leu Leu Ala Xaa Ser Leu Ile Leu Ser Pro Ser Pro Arg Pro Val Leu  
 -15 -10 -5  
 ggc ttt ttc aga caa gtg cat ctc cta acc agg tca cat ttc agc cgc 207  
 Gly Phe Phe Arg Gln Val His Leu Leu Thr Arg Ser His Phe Ser Arg  
 1 5 10 15  
 tgg g 211  
 Trp

<210> 247  
 <211> 359  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 249..359

<221> sig\_peptide  
 <222> 249..308  
 <223> Von Heijne matrix  
 score 6.19999980926514  
 seq LLFICPPPPPIISA/SS

<400> 247  
 tttcagaatt ttgtgcagga atatctgagt atttctaatt agattagaat gtcagaatac 60  
 attcatggac atatgagggg ttttttaaaa ttttttttag atacccttca ccttgaacat 120  
 ttattatttc ttgtgttgga gaacaatcca aatctctcct agatgttttg aaatgtgcaa 180  
 tgtattgtta gctgtagtca ccctactgtg ctattgaata ctagagcttg ttccttctgt 240  
 ctaactgt atg att ata ctc att aac caa ctt ctc ttc atc tgt ccc cca 290  
 Met Ile Ile Leu Ile Asn Gln Leu Leu Phe Ile Cys Pro Pro  
 -20 -15 -10  
 cct cca ccc atc tca gcc tct agt aac tac cat ttt act ctc tac ctc 338  
 Pro Pro Pro Ile Ser Ala Ser Ser Asn Tyr His Phe Thr Leu Tyr Leu  
 -5 1 5 10  
 cat gac att aac ttt ttt agc 359  
 His Asp Ile Asn Phe Phe Ser  
 15

<210> 248  
 <211> 236  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 182..235

<221> sig\_peptide  
 <222> 182..226  
 <223> Von Heijne matrix  
 score 6.19999980926514  
 seq DVLLQLLLRVCSF/RT

<400> 248  
 attgggttaa tttcactgca ctgactatatt ttagatatat attctttgtg ccttcactag 60  
 aactcctctt acttcatgat atcttaacta taaaatcatc caaccatgaa aacaagcaca 120  
 caagaaacag aaacaaaaa gtcacaaaaa agcataaaact gttagcattg atccatgatg 180  
 a atg act gat gta tta ctt caa ttg cta tta aga gtg tgt tct ccc agg 229  
 Met Thr Asp Val Leu Leu Gln Leu Leu Leu Arg Val Cys Ser Pro Arg  
 -15 -10 -5 1  
 acc agg g 236  
 Thr Arg

<210> 249  
 <211> 342  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 266..340

<221> sig\_peptide  
 <222> 266..304  
 <223> Von Heijne matrix  
 score 6.19999980926514  
 seq MGLFLCCSLIFC/LV

<400> 249  
 taggctatatt cttaattttc cttctaggat tcttatagtt tgaagtttta catttagatc 60  
 gtcaatccat cttgagttca tttttgtata tgatgaaaag taggggtctg attttattct 120  
 tctgcataag accagttatc ccagaaccgt ttgttgaata ggaagttctt ttctcattgc 180  
 ttgtttgtgg ggactttgtc aaagatcaaa tagttatagg tgtgtggctg tatttcaggg 240  
 tttctttatt ccatttcact gatct atg ggt ctg ttt ttg tgc tgc tct tta 292  
 Met Gly Leu Phe Leu Cys Cys Ser Leu  
 -10 -5  
 ctg ata ttc tgt ctg gtt gtt cta atc ata act gaa ctg ggc tat ggg 340  
 Leu Ile Phe Cys Leu Val Val Leu Ile Ile Thr Glu Leu Gly Tyr Gly  
 1 5 10  
 gg 342

<210> 250  
 <211> 382  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 291..380

<221> sig\_peptide  
 <222> 291..332  
 <223> Von Heijne matrix  
 score 6.19999980926514  
 seq GSWALTWLHPAEA/GT

<221> misc\_feature  
 <222> 264..265,279..280  
 <223> n=a, g, c or t

<400> 250  
 atacagcggc ctctgacacc agcacagcaa acccgccggg atcaaagtgt accagtcggc 60  
 agcatgggta aggagagggg tttccaatca cccattgcct gctctgtctg cccctaattt 120  
 ggaaaggccc tctccagaa aatgctagaa aacctgagt gggagctggg gagggagtag 180  
 tggactctgc ttcattgtcc ccagtctgca caccctctcc cccaccacc cactgcattt 240  
 cccagctcag ccaaactttc tgannaagac gggcagagnn ctgctgggag atg gga 296  
 Met Gly  
 tcc tgg gcc ctg act tgg ctc cat cca gca gag gct ggg acc agg gtg 344  
 Ser Trp Ala Leu Thr Trp Leu His Pro Ala Glu Ala Gly Thr Arg Val  
 -10 -5 1  
 cct ttc tgc agc tgg gaa aaa tca gat ggg cgc tct ta 382  
 Pro Phe Cys Ser Trp Glu Lys Ser Asp Gly Arg Ser  
 5 10 15

<210> 251



<211> 303  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 108..302

<221> sig\_peptide  
 <222> 108..233  
 <223> Von Heijne matrix  
 score 6.19999980926514  
 seq LSVLSLVINFSWS/RK

<221> misc\_feature  
 <222> 279  
 <223> n=a, g, c or t

<400> 251  
 aaaagctgtg aggttgtaac tcagttcagt agtatttata aatatttggt ttccactttt 60  
 gtgcatatta tacaaatgat ggatataaaa ttgttttwa ccatwta atg atg ctt 116  
 Met Met Leu  
 -40  
 rmw wwr rra aga gga tat cct cat aga act gaa cgt tat gat gga ttt 164  
 Xaa Xaa Xaa Arg Gly Tyr Pro His Arg Thr Glu Arg Tyr Asp Gly Phe  
 -35 -30 -25  
 tta aaa tat tct gac cca aat gat att gca ttg tca gta ctg tcc ctg 212  
 Leu Lys Tyr Ser Asp Pro Asn Asp Ile Ala Leu Ser Val Leu Ser Leu  
 -20 -15 -10  
 gtt att aat ttc tcc tgg agt aga aaa tgc ttt gtt cct tac tat atc 260  
 Val Ile Asn Phe Ser Trp Ser Arg Lys Cys Phe Val Pro Tyr Tyr Ile  
 -5 1 5  
 cca ttt aaa cct tac cgv nta cct tac ccc acc gcg gcc cgg g 303  
 Pro Phe Lys Pro Tyr Arg Xaa Pro Tyr Pro Thr Ala Ala Arg  
 10 15 20

<210> 252  
 <211> 259  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 106..258

<221> sig\_peptide  
 <222> 106..222  
 <223> Von Heijne matrix  
 score 6.19999980926514  
 seq CFVCXLFVFLLSG/LN

<221> misc\_feature  
 <222> 134  
 <223> n=a, g, c or t

<400> 252  
 attttaaagg atttttttaa ggacctctat agttataagt cagcttaatt aaaaatggat 60  
 attccatagt catatttata tatatatata cacacatata tatgt atg tat gtg tgt 117  
 Met Tyr Val Cys  
 ata tat ata trt tta ana gac ctg tat gat ttt ttt ctt ctt gga act 165  
 Ile Tyr Ile Xaa Leu Xaa Asp Leu Tyr Asp Phe Phe Leu Leu Gly Thr

[illegible]

145

Phe Cys Ile Phe Phe Ser Val Pro Ala Leu Gln Pro Arg Arg Leu  
 1 5 10

<210> 255  
 <211> 320  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 233..319

<221> sig\_peptide  
 <222> 233..310  
 <223> Von Heijne matrix  
 score 6.09999990463257  
 seq MLTLLGFPSKALT/FI

<221> misc\_feature  
 <222> 129  
 <223> n=a, g, c or t

<400> 255  
 caagttgtct cctgcgtagt gtctattagc tcttgaattt cttcaagatc catatactga 60  
 aacacttcac tctccaactt ttttgccata ttgacaatca ctttcataat ttcacttatt 120  
 gacyctgynw haaatcmtgt gaagyhatgc agahcatctg gacacagctt tctccagcag 180  
 ggatyyatdg ttttgggctt gaagggggtt cacggctttt tctataacaa cg atg gca 238  
 Met Ala  
 -25  
 tct tca atg ctg waa tcc ttc cag act ttc atg atg ttg act cta ttg 286  
 Ser Ser Met Leu Xaa Ser Phe Gln Thr Phe Met Met Leu Thr Leu Leu  
 -20 -15 -10  
 ggt ttc cct tcc aaa gct ttg aca ttc att tcc a 320  
 Gly Phe Pro Ser Lys Ala Leu Thr Phe Ile Ser  
 -5 1

<210> 256  
 <211> 305  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 205..303

<221> sig\_peptide  
 <222> 205..264  
 <223> Von Heijne matrix  
 score 6.09999990463257  
 seq LLSLPGSFIPGNC/RP

<400> 256  
 tttgttttat ttggttatatt gttttgtttt gtttctctga ggccaatggg tgggaggaag 60  
 tataaagaag tgtaaacagg aaagccagct gggcctggag ttccaagtgc ccatatttca 120  
 tcagcttcct ctccataact gtggcaggga cacttaaccc ttccctggct gtgagaagtt 180  
 attctctgag ggctggtgag caga atg gga aga tct aag agg cag ctg ctt 231  
 Met Gly Arg Ser Lys Arg Gln Leu Leu  
 -20 -15  
 tcc ttg cct ggt tcc ttt atc cct ggg aat tgc agg cca agg att ctg 279  
 Ser Leu Pro Gly Ser Phe Ile Pro Gly Asn Cys Arg Pro Arg Ile Leu  
 -10 -5 1 5  
 agc aat ggw gaa gwc aga agg aag gg 305

146

Ser Asn Gly Glu Xaa Arg Arg Lys  
10

<210> 257  
<211> 181  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 37..180

<221> sig\_peptide  
<222> 37..111  
<223> Von Heijne matrix  
score 6.09999990463257  
seq CFFLIIFLLPPLPA/MI

<400> 257  
tttctaattgc tattatcctg mtagtgamta agtctc atg aga tct gat ggg ttt 54  
Met Arg Ser Asp Gly Phe  
-25 -20  
atc agg ggt ttc tgc ttc tgc ttc ttc cta att ttt ctc ctg cca ccg 102  
Ile Arg Gly Phe Cys Phe Cys Phe Phe Leu Ile Phe Leu Leu Pro Pro  
-15 -10 -5  
ctt cct gcc atg ata ctg agg cct ctg cag cca tgt gga att ata agt 150  
Leu Pro Ala Met Ile Leu Arg Pro Leu Gln Pro Cys Gly Ile Ile Ser  
1 5 10  
cca att aaa cct ctt ttt cct ttt ttt ttt t 181  
Pro Ile Lys Pro Leu Phe Pro Phe Phe Phe  
15 20

<210> 258  
<211> 236  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 119..235

<221> sig\_peptide  
<222> 119..166  
<223> Von Heijne matrix  
score 6.09999990463257  
seq LWTASSLPLSTHS/QR

<400> 258  
caaaaaaatc agtctttaag catttgcttg gtaaggtttc ttaagattag gtttataata 60  
caaccatctg taatgtatct stcgtttgag cttgtggggcc atacaattca ttaactag 118  
atg aat aca ttg tgg aca gca tcc tca cta ccc ctc tct act cac tca 166  
Met Asn Thr Leu Trp Thr Ala Ser Ser Leu Pro Leu Ser Thr His Ser  
-15 -10 -5  
caa aga acc atg ata cac tgg aat gtt ttt ctc tgg aat tct ttc tac 214  
Gln Arg Thr Met Ile His Trp Asn Val Phe Leu Trp Asn Ser Phe Tyr  
1 5 10 15  
tct tgt att aaa att ttt ccc c 236  
Ser Cys Ile Lys Ile Phe Pro  
20

<210> 259  
<211> 265  
<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 128..265

<221> sig\_peptide

<222> 128..220

<223> Von Heijne matrix

score 6.09999990463257

seq CLIGLLVPLLGWG/NQ

<400> 259

gacttaggat	ttgagcatct	ttctgttatg	ctggtgcccc	actcctattg	caatactccc	60
cttcttaaga	aagtttttct	agactaatgt	ctagattaaa	cttcttttct	ttgacaataa	120
tgatgcc	atg act tgg	aca aaa tgc	cca ttg cct	ctg ggt cct	gct ttc	169
	Met Thr Trp Thr	Lys Cys Pro Leu	Pro Leu Gly	Pro Ala Phe		
	-30	-25	-20			
ttc acc	cag tgc tgc	ctt att gga	ctc ctt gtg	cct ctc ctt	ggc tgg	217
Phe Thr	Gln Cys Cys	Leu Ile Gly	Leu Leu Val	Pro Leu Leu	Gly Trp	
	-15	-10	-5			
gga aat	cag aat aca	cag tgg tat	ccc act tct	aag atg cct	gat ggg	265
Gly Asn	Gln Asn Thr	Gln Trp Tyr	Pro Thr Ser	Lys Met Pro	Asp Gly	
1	5		10		15	

<210> 260

<211> 272

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 162..272

<221> sig\_peptide

<222> 162..257

<223> Von Heijne matrix

score 6.09999990463257

seq IVYFLVLLRVLYT/LQ

<400> 260

cacaagggtg	attttaaatt	cttaaaaaat	ttttcaaaat	ctttccaaat	gaaacaagat	60
ttattgttaa	tctacagaaa	tatcctccat	tcactttgat	attttaaata	catcgatcat	120
tttaggtaga	gcatttttat	gaccactcat	tgcttagtct	g atg ggg	agg agc aat	176
				Met Gly Arg	Ser Asn	
				-30		
gat ttt	agg ttt gcc	ttt cta aca	tgc ttt ctt	gga tgg	gaa ata gta	224
Asp Phe	Arg Phe Ala	Phe Leu Thr	Cys Phe Leu	Gly Trp Glu	Ile Val	
	-25	-20	-15			
tat ttc	ttg gtg ctt	ctt cgt gtt	tta tac act	tta caa	tgg ggt ggg	272
Tyr Phe	Leu Val Leu	Leu Arg Val	Leu Tyr Thr	Leu Gln	Trp Gly Gly	
	-10	-5	1		5	

<210> 261

<211> 98

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 26..97

<221> sig\_peptide

&lt;222&gt; 26..79

&lt;223&gt; Von Heijne matrix

score 6.09999990463257

seq LTSFLTYMPLISS/SC

&lt;400&gt; 261

```

tttctaggta tacaatcata ttata atg aaa aca gat aat ttg act tct ttt      52
                               Met Lys Thr Asp Asn Leu Thr Ser Phe
                               -15                               -10
ctt aca tat atg cct ctt att tct tcc tct tgc tca att gct ccc t      98
Leu Thr Tyr Met Pro Leu Ile Ser Ser Ser Cys Ser Ile Ala Pro
                               -5                               1                               5

```

&lt;210&gt; 262

&lt;211&gt; 419

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 28..417

&lt;221&gt; sig\_peptide

&lt;222&gt; 28..264

&lt;223&gt; Von Heijne matrix

score 6.09999990463257

seq ATVAVLSFILSSA/AK

&lt;400&gt; 262

```

attccccgggc cctggcttct tggcgcg atg agg ttc cgg ttc tgt ggt gat ctg      54
                               Met Arg Phe Arg Phe Cys Gly Asp Leu
                               -75
gac tgt ccc gac tgg gtc ctg gca gaa atc agc acg ctg gcc aag atg      102
Asp Cys Pro Asp Trp Val Leu Ala Glu Ile Ser Thr Leu Ala Lys Met
-70                               -65                               -60                               -55
tcc tct gtg aag ttg cgg ctg ctc tgc agc cag gta cta aag gag ctg      150
Ser Ser Val Lys Leu Arg Leu Leu Cys Ser Gln Val Leu Lys Glu Leu
                               -50                               -45                               -40
ctg gga cag ggg att gat tat gag aag atc ctg aag ctc acg gct gac      198
Leu Gly Gln Gly Ile Asp Tyr Glu Lys Ile Leu Lys Leu Thr Ala Asp
                               -35                               -30                               -25
gcc aag ttt gag tca ggc gat gtg aag gcc aca gtg gca gtg ctg agt      246
Ala Lys Phe Glu Ser Gly Asp Val Lys Ala Thr Val Ala Val Leu Ser
                               -20                               -15                               -10
ttc atc ctc tcc agt gcg gcc aag cac agt gtc gat ggc gaa tcc ttg      294
Phe Ile Leu Ser Ser Ala Ala Lys His Ser Val Asp Gly Glu Ser Leu
                               -5                               1                               5                               10
tcc agt gaa ctg cag cag ctg ggg ctg ccc aaa gag cac gcg gcc agc      342
Ser Ser Glu Leu Gln Gln Leu Gly Leu Pro Lys Glu His Ala Ala Ser
                               15                               20                               25
ctg tgc cgc tgt tat gag gag aag caa agc ccc ttg cag aag cac ttg      390
Leu Cys Arg Cys Tyr Glu Glu Lys Gln Ser Pro Leu Gln Lys His Leu
                               30                               35                               40
cgg gtc tgc agc cta cgc atg aat agg tt      419
Arg Val Cys Ser Leu Arg Met Asn Arg
                               45                               50

```

&lt;210&gt; 263

&lt;211&gt; 371

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 133..369

&lt;221&gt; sig\_peptide

&lt;222&gt; 133..174

&lt;223&gt; Von Heijne matrix

score 6.09999990463257

seq FLAALFTVAKIWK/QP

&lt;400&gt; 263

cactatggag aactgtatgg cgggtcctca aaaaactaaa aatagaactc ccatatgatc 60

cagcaatccc attgctaggt atataccccc ccaaaaaagg aaatcagtat atgaaagaga 120

tatctgaatc cc atg ttt ctt gca gca ctg ttt aca gta gct aag att tgg 171

Met Phe Leu Ala Ala Leu Phe Thr Val Ala Lys Ile Trp

-10

-5

aag caa cct aag tgt tca tca aca aac aaa tgg aca aag aaa atg tgg 219

Lys Gln Pro Lys Cys Ser Ser Thr Asn Lys Trp Thr Lys Lys Met Trp

1

5

10

15

tac ata tac aca atg gag tac tat tca gcc ata aaa aaa gat gat atc 267

Tyr Ile Tyr Thr Met Glu Tyr Tyr Ser Ala Ile Lys Lys Asp Asp Ile

20

25

30

ctg tca ttt gca aca ata tgg atg gaa ctg gag agc att aca tta agt 315

Leu Ser Phe Ala Thr Ile Trp Met Glu Leu Glu Ser Ile Thr Leu Ser

35

40

45

gaa ata agt ggg sca cca aaa gac aaa ctt ctc atg ttc tca ctc att 363

Glu Ile Ser Gly Xaa Pro Lys Asp Lys Leu Leu Met Phe Ser Leu Ile

50

55

60

tgt gga ag 371

Cys Gly

65

&lt;210&gt; 264

&lt;211&gt; 283

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 194..283

&lt;221&gt; sig\_peptide

&lt;222&gt; 194..274

&lt;223&gt; Von Heijne matrix

score 6.09999990463257

seq LSILQSLVPAAGA/XS

&lt;400&gt; 264

ctcattccct gtccctcgat cacagtctct tctcactaca gtgtcgccgc ctctgcctgc 60

gtascccggc catggctctg tagcctcgac ccctttgtgc ccccggcccg tctccgcgct 120

caccacgcct gcgctctccg ctcccacctt ctttcttcag ccgaggccgc cgccgcctct 180

ccttgctgca gcc atg gag tct tcc act ttc gcc ttg gtg cct gtc ttc 229

Met Glu Ser Ser Thr Phe Ala Leu Val Pro Val Phe

-25

-20

gcc cac ctg agc atc ctc cag agc ctc gtg cca gct gct ggt gca gyc 277

Ala His Leu Ser Ile Leu Gln Ser Leu Val Pro Ala Ala Gly Ala Xaa

-15

-10

-5

1

tct cct 283

Ser Pro

&lt;210&gt; 265

&lt;211&gt; 370

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 117..368

&lt;221&gt; sig\_peptide

&lt;222&gt; 117..350

&lt;223&gt; Von Heijne matrix

score 6.09999990463257

seq LLWFLQTFFGIA/SL

&lt;400&gt; 265

```

aaagcgcgct cccggggagg tgttcagacc atggctacgg cagccggcgc gacctacttt      60
cagcgaggca gtctgttctg gttcacagtc atcaccctca gctttggcta ctacac atg      119
                                         Met

```

```

ggg tgt ctt ctg gcc tca gag tat ccc tta tca gaa cct tgg gcc cct      167
Gly Cys Leu Leu Ala Ser Glu Tyr Pro Leu Ser Glu Pro Trp Ala Pro
      -75                      -70                      -65

```

```

ggg ccc ttc act cag tac ttg gtg gac cac cat cac acc ctc ctg tgc      215
Gly Pro Phe Thr Gln Tyr Leu Val Asp His His His Thr Leu Leu Cys
      -60                      -55                      -50

```

```

aat ggg tat tgg ctt gcc tgg ctg att cat gtg gga gag tcc ttg tat      263
Asn Gly Tyr Trp Leu Ala Trp Leu Ile His Val Gly Glu Ser Leu Tyr
      -45                      -40                      -35                      -30

```

```

gcc ata gta ttg tgc aag cat aaa ggc atc aca agt ggt cgg gct cag      311
Ala Ile Val Leu Cys Lys His Lys Gly Ile Thr Ser Gly Arg Ala Gln
      -25                      -20                      -15

```

```

cta ctc tgg ttc cta cag act ttc ttc ttt ggg ata gcg tct ctc asc      359
Leu Leu Trp Phe Leu Gln Thr Phe Phe Phe Gly Ile Ala Ser Leu Xaa
      -10                      -5                      1

```

```

atc ttg att gc      370
Ile Leu Ile
      5

```

&lt;210&gt; 266

&lt;211&gt; 274

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 178..273

&lt;221&gt; sig\_peptide

&lt;222&gt; 178..225

&lt;223&gt; Von Heijne matrix

score 6.09999990463257

seq WIWVASILLRIFA/SV

&lt;400&gt; 266

```

tatcgtgaaa gaatattgaa ttttatcaaa tctttttttg tatctgttga gatgattaca      60
tgggtattat ccttcattct gttgatgtga tgtatcacat ttattgattt gcatatgttg      120
aaccctcctt gcatccctgg aatgattcct acttcattat agtgataat ctttttg      177
atg tgc tgt tgg att tgg gtt gct agt att ttg ttg aga att ttt gca      225
Met Cys Cys Trp Ile Trp Val Ala Ser Ile Leu Leu Arg Ile Phe Ala
      -15                      -10                      -5

```

```

tct gtg tta atc agg gat att tac ctg tgg ttt tct ttt ttt ttt t      274
Ser Val Leu Ile Arg Asp Ile Tyr Leu Trp Phe Ser Phe Phe Phe Phe
      1                      5                      10                      15

```

&lt;210&gt; 267

&lt;211&gt; 342

&lt;212&gt; DNA



<213> Homo sapiens

<220>

<221> CDS

<222> 232..342

<221> sig\_peptide

<222> 232..300

<223> Von Heijne matrix

score 6.09999990463257

seq LLFFXLWLRYKES/GR

<221> misc\_feature

<222> 158

<223> n=a, g, c or t

<400> 267

```

caagttatct caatttcttc tgagaagaaa tatagtttca aaatcaatca ataaagataa      60
tcctctgata aagtaagatc tgaatataca aatcatgggtt acagtaatct taccattata      120
tataaattac ctctcaaaca aatggggccat tcagaarnrg gctcagagtg aattagctgg      180
aggggttgctc aagggtcata gtttttactg ctttgaagag attatcactg g atg att      237
                                   Met Ile
tcc tca cat tta tat aac ttc agt ctc ctg ttc ttt kta ctc tgg ctg      285
Ser Ser His Leu Tyr Asn Phe Ser Leu Leu Phe Phe Xaa Leu Trp Leu
   -20                               -15               -10
agg tac aag gaa tca gga aga gag ggc aac tgt gag gaa gga gca ttc      333
Arg Tyr Lys Glu Ser Gly Arg Glu Gly Asn Cys Glu Glu Gly Ala Phe
   -5                               1               5               10
tcc agg tgg      342
Ser Arg Trp

```

<210> 268

<211> 427

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 62..427

<221> sig\_peptide

<222> 62..112

<223> Von Heijne matrix

score 6.09999990463257

seq RLLLLPRPPASTG/AS

<400> 268

```

agttgagtgg aaatgggcaa cggcgggagg agcggcctgc agcaggggaa ggggaacgtg      60
g atg ggg tgg cag cga ctc cta ctg ctg cct cgg cct cct gcc agt aca      109
  Met Gly Trp Gln Arg Leu Leu Leu Leu Pro Arg Pro Pro Ala Ser Thr
    -15                               -10               -5
ggt gca tcg aat gca acc agg rrg cca aag agk ttg tac cga grc tat      157
Gly Ala Ser Asn Ala Thr Arg Xaa Pro Lys Xaa Leu Tyr Arg Xaa Tyr
   1                               5               10               15
aac cac ggt gtg ctg aag ata acc atc tgt aaa tcc tgc cag aaa cct      205
Asn His Gly Val Leu Lys Ile Thr Ile Cys Lys Ser Cys Gln Lys Pro
    20                               25               30
gta gac aaa tat atc gag tat gat cct gtt atc atc ttg awk aat gct      253
Val Asp Lys Tyr Ile Glu Tyr Asp Pro Val Ile Ile Leu Xaa Asn Ala
    35                               40               45
ata ttg tgc aaa gct cad gcc tac agr cat att ctt ttc aat act caa      301
Ile Leu Cys Lys Ala Xaa Ala Tyr Arg His Ile Leu Phe Asn Thr Gln

```

152

50	55	60	
ata aat aac aaa ctg cct att tta ttg gca ttt tta cct tcc tgt ggv			349
Ile Asn Asn Lys Leu Pro Ile Leu Leu Ala Phe Leu Pro Ser Cys Gly			
65	70	75	
dga acg gcc cat gac ggc aaa aaa aag ccc aac ttc att ttg ctg ctg			397
Xaa Thr Ala His Asp Gly Lys Lys Lys Pro Asn Phe Ile Leu Leu Leu			
80	85	90	95
aaa sat tat tat tat cta gct acg gaa aac			427
Lys Xaa Tyr Tyr Tyr Leu Ala Thr Glu Asn			
100	105		

<210> 269  
 <211> 143  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 20..142  
 <221> sig\_peptide  
 <222> 20..76  
 <223> Von Heijne matrix  
 score 6  
 seq LLALVVRVILSTA/IL

<400> 269	
ctttctttgc ggaatcacc atg gcg gct ggg gta agt ttg ctg gct ctg gtg	52
Met Ala Ala Gly Val Ser Leu Leu Ala Leu Val	
-15	-10
gtt cgg gtc atc cta tcc acc gcc atc ctt tgc ccg agt ggg gcc agt	100
Val Arg Val Ile Leu Ser Thr Ala Ile Leu Cys Pro Ser Gly Ala Ser	
-5	1
cgg cgc cag agg agt tct gag gtt gag tgg gga act gat tcg g	143
Arg Arg Gln Arg Ser Ser Glu Val Glu Trp Gly Thr Asp Ser	
10	15
	20

<210> 270  
 <211> 79  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 23..79  
 <221> sig\_peptide  
 <222> 23..67  
 <223> Von Heijne matrix  
 score 6  
 seq PLFWLILCSGLLC/NK

<221> misc\_feature  
 <222> 2..3  
 <223> n=a, g, c or t

<400> 270	
tnnngctaadc ttgcttggtac tt atg aat cct tta ttc tgg ttg att ctc tgc	52
Met Asn Pro Leu Phe Trp Leu Ile Leu Cys	
-15	-10
tct ggg tta tta tgt aac aag tca ttt	79
Ser Gly Leu Leu Cys Asn Lys Ser Phe	

-5 1

<210> 271  
 <211> 121  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 61..120

<221> sig\_peptide  
 <222> 61..114  
 <223> Von Heijne matrix  
 score 6  
 seq ISIFLSSLSSLS/LF

<400> 271  
 cttccttaag aagcggtttc tctccctct tttctctctc tcacctgggt ttgtttgtcc 60  
 atg aga ggg gct tgg ata agt ata ttt ctt tct tct cta tct ctc tct 108  
 Met Arg Gly Ala Trp Ile Ser Ile Phe Leu Ser Ser Leu Ser Leu Ser  
 -15 -10 -5  
 ctc tct ctt ttt t 121  
 Leu Ser Leu Phe  
 1

<210> 272  
 <211> 292  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 196..291

<221> sig\_peptide  
 <222> 196..267  
 <223> Von Heijne matrix  
 score 6  
 seq LFVLLPHFFLSFL/SP

<400> 272  
 ctcattggact gtggctgtct tattttatgt ctctaatacc agattatgaa aatcacagaa 60  
 aaaaggaaaa aatattatctt ccaaagagta agttatgaag ccatgtaga aacctatg 120  
 acaatatgaa tttcttttat ctgtcaatct caaggtagaa ttcctcatat ttctgataat 180  
 gccaaatacc atgaa atg tct caa aaa aga ctt gac ttt ata tac cag ttg 231  
 Met Ser Gln Lys Arg Leu Asp Phe Ile Tyr Gln Leu  
 -20 -15  
 ttt gtc ttg ctg cct cac ttc ttc ctt tct ttt ctt tct ccc ttt tat 279  
 Phe Val Leu Leu Pro His Phe Phe Leu Ser Phe Leu Ser Pro Phe Tyr  
 -10 -5 1  
 ctg cac cca tgg g 292  
 Leu His Pro Trp  
 5

<210> 273  
 <211> 158  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 2..157

```

<221> sig_peptide
<222> 2..100
<223> Von Heijne matrix
      score 6
      seq LAHFLIGLTVCFG/EG

<400> 273
c atg tac ctg tac ctg ttg tcc att tgt atg tct tct ttg aag aaa tgt      49
  Met Tyr Leu Tyr Leu Leu Ser Ile Cys Met Ser Ser Leu Lys Lys Cys
      -30                -25                -20
cta ttc aag ttc tta gcc cac ttt tta atc ggg tta aca gtt tgt ttt      97
Leu Phe Lys Phe Leu Ala His Phe Leu Ile Gly Leu Thr Val Cys Phe
      -15                -10                -5
ggg gag ggr wgg cta atg agt tat agg agt tct tat tta tta ctt aaa      145
Gly Glu Gly Xaa Leu Met Ser Tyr Arg Ser Ser Tyr Leu Leu Leu Lys
      1                5                10                15
gga cca ccg ggg g
Gly Pro Pro Gly
      158

<210> 274
<211> 113
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 31..111

<221> sig_peptide
<222> 31..96
<223> Von Heijne matrix
      score 6
      seq CLLVFLLTEWTSS/KL

<400> 274
ccttttgtct ttgatgatgg tgacatacag atg ggg ttt tgg tgt gaa tgt cct      54
                        Met Gly Phe Trp Cys Glu Cys Pro
                        -20                -15
ttc tgt ttg tta gtt ttc ctt cta aca gag tgg acc tct agc aaa ctc      102
Phe Cys Leu Leu Val Phe Leu Leu Thr Glu Trp Thr Ser Ser Lys Leu
      -10                -5                1
caa aag acg gg
Gln Lys Thr
      5
      113

<210> 275
<211> 254
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 182..253

<221> sig_peptide
<222> 182..247
<223> Von Heijne matrix
      score 6
      seq VLHLFPLTPASTG/HW

<400> 275
cccacatgcc tgctcccact aggggtctga gtagcaggca ccgaagaagt gagccacgcc      60

```

155

```

ctcttcacac accctttgag gaggacaagg gaacttttcc tgtttcagaa agttgtgttg 120
agaagaatgg caaggctaac agggcaggtg tccgggcgga ggggcggaac tggctgttgg 180
c atg tgg tgg ggg aga tgc ttc atc cgg gtc ttg cat ttg ttc cct ctg 229
  Met Trp Trp Gly Arg Cys Phe Ile Arg Val Leu His Leu Phe Pro Leu
    -20 -15 -10
aca cca gcc tcg aca gga cac tgg g 254
Thr Pro Ala Ser Thr Gly His Trp
  -5 1

```

<210> 276  
 <211> 364  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 189..362

<221> sig\_peptide  
 <222> 189..275  
 <223> Von Heijne matrix  
 score 6  
 seq LFMALPPVLSSHG/SR

```

<400> 276
acttgcaactt gcttgttggg gtcagagccc gtccctaaac cagggctcca tatgggctgc 60
ctgtctgccg caacacagcc tagcggggaa acagtagaaa tgccacttct atgtatttat 120
catatttatt ttgagataat taacgaagac gttaaataaa gccagactgc actgacctct 180
ggggcgcc atg cga gac ccc ctc gcg gac atg gta cac agt tat tta tca 230
  Met Arg Asp Pro Leu Ala Asp Met Val His Ser Tyr Leu Ser
    -25 -20
tcg tct ttg ttc atg gcc ctt cca cca gtg ctg agc tca cat ggc agc 278
Ser Ser Leu Phe Met Ala Leu Pro Pro Val Leu Ser Ser His Gly Ser
-15 -10 -5 1
agg aac ctg aga atc tgg ggg agt cca ttt ggt gga gcg ctg act aag 326
Arg Asn Leu Arg Ile Trp Gly Ser Pro Phe Gly Gly Ala Leu Thr Lys
  5 10 15
ggc aaa gca ccc cca acc cca gca caa cca gcc ctg gg 364
Gly Lys Ala Pro Pro Thr Pro Ala Gln Pro Ala Leu
  20 25

```

<210> 277  
 <211> 130  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 46..129

<221> sig\_peptide  
 <222> 46..96  
 <223> Von Heijne matrix  
 score 6  
 seq WLCLPCSLCVSQL/LP

```

<400> 277
gtctttgcag gmvgtgttgt gctccaacag ggagctgagt ttgtc atg agc agt gcc 57
  Met Ser Ser Ala
    -15
tgg ctg tgt ctg cca tgc tcc ctg tgt gtg tcc cag ctc ctt ccc tct 105
Trp Leu Cys Leu Pro Cys Ser Leu Cys Val Ser Gln Leu Leu Pro Ser
  -10 -5 1

```

156

tat tcc ctg ttg atc cca gcc ccg g  
 Tyr Ser Leu Leu Ile Pro Ala Pro  
           5                          10

130

<210> 278  
 <211> 184  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 103..183

<221> sig\_peptide  
 <222> 103..165  
 <223> Von Heijne matrix  
       score 5.90000009536743  
       seq LSLLGPLXPPMRA/CS

<400> 278  
 cattatgttg acatttctag ctacaaggcc agtattttac aaaataaggc cttttccctt       60  
 aattaagggt gtgacagata aaagtatat cccagctgac tc atg tca ccc atg       114  
   Met Ser Pro Met  
   -20  
 tgg gca ggc cta tta tcc cta ctt ggc ccg ctc wgt ccg cct atg agg       162  
 Trp Ala Gly Leu Leu Ser Leu Leu Gly Pro Leu Xaa Pro Pro Met Arg  
           -15  -10  -5  
 gct tgc tct gtg tgc gta ctc t       184  
 Ala Cys Ser Val Cys Val Leu  
           1  5

<210> 279  
 <211> 265  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 149..265

<221> sig\_peptide  
 <222> 149..202  
 <223> Von Heijne matrix  
       score 5.90000009536743  
       seq LSIADLLPSSSFA/NP

<400> 279  
 tgagcatttc ccaacatatac tgttgagttt ttaactcttt ttatgatctt ttttatttct       60  
 aggagttcta tttagttctt ttccaagtca gctatgtcat gttttaawag tttccwgtcc       120  
 ttdgcrtrwc twttctawct tgagggtt atg tct tta aac gag tta agc ata       172  
   Met Ser Leu Asn Glu Leu Ser Ile  
   -15  
 gct gat tta tta ccc agc tca tcc ttt gct aat ccc aag ctg agt ggg       220  
 Ala Asp Leu Leu Pro Ser Ser Ser Phe Ala Asn Pro Lys Leu Ser Gly  
           -10  -5  1  5  
 ccg att tct atc tcg gtc act tca gct ggt tct cct ccc ggg gcg       265  
 Pro Ile Ser Ile Ser Val Thr Ser Ala Gly Ser Pro Pro Gly Ala  
           10  15  20

<210> 280  
 <211> 188  
 <212> DNA  
 <213> Homo sapiens

```

<220>
<221> CDS
<222> 110..187

<221> sig_peptide
<222> 110..154
<223> Von Heijne matrix
      score 5.90000009536743
      seq DLLGTAFLEGLA/AY

<400> 280
taataataat aataataaat ttttctgtta gattagtaga tgtaagatt atggacaaaa      60
tccaacgtag attggagtat agagaagtgg acctttctgt gtggtcttg atg aaa gac      118
                                     Met Lys Asp
                                     -15
tta ctt ggc act gcc ttt ctg gag gga agt tta gca gca tat ctc acc      166
Leu Leu Gly Thr Ala Phe Leu Glu Gly Ser Leu Ala Ala Tyr Leu Thr
      -10      -5      1
atg gcc aat ata acc cat gtg g      188
Met Ala Asn Ile Thr His Val
      5      10

<210> 281
<211> 177
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 91..177

<221> sig_peptide
<222> 91..147
<223> Von Heijne matrix
      score 5.90000009536743
      seq HLFMCLLTICISS/LE

<400> 281
gccaaacttgt ttttctgtgt agttttgatt agagccattc tactgggtgt gaagtgatat      60
tttgatgttg ttttgatttg catttccttg atg gct aat gac att aag cat ctt      114
                                     Met Ala Asn Asp Ile Lys His Leu
                                     -15
ttc atg tgc tta ttg acc ata tgt ata tct tct ttg gag aaa ctt cca      162
Phe Met Cys Leu Leu Thr Ile Cys Ile Ser Ser Leu Glu Lys Leu Pro
      -10      -5      1      5
ttc ttt ttt ttt ttt      177
Phe Phe Phe Phe Phe
      10

<210> 282
<211> 336
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 42..335

<221> sig_peptide
<222> 42..113
<223> Von Heijne matrix
      score 5.90000009536743

```

158

seq ATLVGFTVGSVLG/QI

&lt;400&gt; 282

```

ttgcctatta ctcttatatc tacagtgtgg tggacctggg c atg tac cag aaa gtc      56
                                     Met Tyr Gln Lys Val
                                     -20
aca agt tac tgt cga agt gcc act ttg gtg ggc ttt aca gtg ggc tct      104
Thr Ser Tyr Cys Arg Ser Ala Thr Leu Val Gly Phe Thr Val Gly Ser
          -15          -10          -5
gtc cta ggg caa atc ctt gtc tca gtg gca ggc tgg tcg ctg ttc agc      152
Val Leu Gly Gln Ile Leu Val Ser Val Ala Gly Trp Ser Leu Phe Ser
          1          5          10
ctg aat gtc atc tct ctt acc tgt gtt tca gtg gct ttt gct gtg gcc      200
Leu Asn Val Ile Ser Leu Thr Cys Val Ser Val Ala Phe Ala Val Ala
          15          20          25
tgg ttt tta cct atg cca cag aag agc ctc ttc ttt cac cac att cct      248
Trp Phe Leu Pro Met Pro Gln Lys Ser Leu Phe Phe His His Ile Pro
          30          35          40          45
tct acc tgc cag aga gtg aat ggc atc aag gta caa aat ggt ggc att      296
Ser Thr Cys Gln Arg Val Asn Gly Ile Lys Val Gln Asn Gly Gly Ile
          50          55          60
gtt act gac acc cag ctt cta aca cct tcc tgg ctg gga g      336
Val Thr Asp Thr Gln Leu Leu Thr Pro Ser Trp Leu Gly
          65          70

```

&lt;210&gt; 283

&lt;211&gt; 294

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 238..294

&lt;221&gt; sig\_peptide

&lt;222&gt; 238..288

&lt;223&gt; Von Heijne matrix

score 5.90000009536743

seq ALFFLLRIAWLLG/LF

&lt;221&gt; misc\_feature

&lt;222&gt; 227

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 283

```

acatacacgt caattaatct gattcatccc ataaacaaaa ctaaagataa aaaccatgtg      60
attatctcaa tatatgcaga aaaggctttc aataaaattc aaggcctctc catattaaaa      120
actctaaaaa atctgggtat tgaggaarca tagctcaaaa gtgatgrgct gttttgtac      180
cagtatcatg ctgttttggt tactgtagcc ctgtagtata gtttgangtt gggtaac      237
atg atg cct cca gct ttg ttc ttt ttg ctg agg att gct tgg cta tta      285
Met Met Pro Pro Ala Leu Phe Phe Leu Leu Arg Ile Ala Trp Leu Leu
          -15          -10          -5
ggg ctc ttt      294
Gly Leu Phe
          1

```

&lt;210&gt; 284

&lt;211&gt; 203

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;



<221> CDS  
 <222> 90..203

<221> sig\_peptide  
 <222> 90..152  
 <223> Von Heijne matrix  
 score 5.90000009536743  
 seq ALSLWASVSPSWM/CR

```

<400> 284
catcttttcgg cctagatgga ggaaccgtgt gctggctggg caggctgctg gcagaggtea      60
ggagggtctct tccctgagcc ctgccatcc atg aac tgt gta act ttg atc cag      113
                               Met Asn Cys Val Thr Leu Ile Gln
                               -20                               -15

gcc ttg tcc ctc tgg gcc tca gtt tcc cca agc tgg atg tgt cgt ccc      161
Ala Leu Ser Leu Trp Ala Ser Val Ser Pro Ser Trp Met Cys Arg Pro
                -10                -5                1

cct gct tca ttc ata atc acc acc acc acc acc acc tgc ggg      203
Pro Ala Ser Phe Ile Ile Thr Thr Thr Thr Thr Thr Cys Gly
      5                10                15
  
```

<210> 285  
 <211> 297  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 240..296

<221> sig\_peptide  
 <222> 240..287  
 <223> Von Heijne matrix  
 score 5.90000009536743  
 seq LLSLMARTDLVFC/SP

<221> misc\_feature  
 <222> 107  
 <223> n=a, g, c or t

```

<400> 285
aggcattgtc taggctgctg ggcacatgag ctccgggatg cccatgtcct ctggccaggc      60
agacacagac ctggggcagc accagctttc tgatggcagc ctgctcnttc caacagttcc      120
ctaccagaat cctgcctcac tggagcagag gatgccagca tcagccggga accactcctg      180
tgctaaaacc gccttggtgg cctgtggctt gaggtcttga tgcggatgaa gccggagga      239
atg ttg tct ctc ctc agt ctc atg gca agg act gat ctt gtt ttc tgt      287
Met Leu Ser Leu Leu Ser Leu Met Ala Arg Thr Asp Leu Val Phe Cys
      -15                -10                -5

tcc cca cgg g      297
Ser Pro Arg
1
  
```

<210> 286  
 <211> 774  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 8..772

<221> sig\_peptide

&lt;222&gt; 8..109

&lt;223&gt; Von Heijne matrix

score 5.90000009536743

seq MAVLAPLIALVYS/VP

&lt;221&gt; misc\_feature

&lt;222&gt; 486,565

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 286

```

agtcggtt atg gtg ggg gag gcg ggg cga gac cta cga cgc cgg cga gca      49
          Met Val Gly Glu Ala Gly Arg Asp Leu Arg Arg Arg Arg Ala
          -30                                -25

gtg gcc gtt acg gcc gaa aag atg gcg gtc ttg gca cct cta att gct      97
Val Ala Val Thr Ala Glu Lys Met Ala Val Leu Ala Pro Leu Ile Ala
-20                                -15                                -10                                -5

ctc gtg tat tcg gtg ccg cga ctt tca cga tgg ctc gcc caa cct tac      145
Leu Val Tyr Ser Val Pro Arg Leu Ser Arg Trp Leu Ala Gln Pro Tyr
          1                                5                                10

tac ctt ctg tcg gcc ctg ctc tct gct gcc ttc cta ctc gtg agg aaa      193
Tyr Leu Leu Ser Ala Leu Leu Ser Ala Ala Phe Leu Leu Val Arg Lys
          15                                20                                25

ctg ccg ccg ctc tgc cac ggt ctg ccc acc caa cgc gaa gac ggt aac      241
Leu Pro Pro Leu Cys His Gly Leu Pro Thr Gln Arg Glu Asp Gly Asn
          30                                35                                40

ccg tgt gac ttt gac tgg aga gaa gtg gag atc ctg atg ttt ctc agt      289
Pro Cys Asp Phe Asp Trp Arg Glu Val Glu Ile Leu Met Phe Leu Ser
          45                                50                                55                                60

gcc att gtg atg atg aag aac cgc aga tcc atc act gtg gag caa cat      337
Ala Ile Val Met Met Lys Asn Arg Arg Ser Ile Thr Val Glu Gln His
          65                                70                                75

ata ggc aac att ttc atg ttt agt aaa gtg gcc aac aca att ctt ttc      385
Ile Gly Asn Ile Phe Met Phe Ser Lys Val Ala Asn Thr Ile Leu Phe
          80                                85                                90

ttc cgc ttg gat att cgc atg ggc cta ctt tac atc aca ctc tgc ata      433
Phe Arg Leu Asp Ile Arg Met Gly Leu Leu Tyr Ile Thr Leu Cys Ile
          95                                100                                105

gtg ttc ctg atg acg tgc aaa ccc ccc cta tat atg ggc cct gag tat      481
Val Phe Leu Met Thr Cys Lys Pro Pro Leu Tyr Met Gly Pro Glu Tyr
          110                                115                                120

atc ang tac ttc aat gat aaa acc att gat gag gaa cta gaa cgg gac      529
Ile Xaa Tyr Phe Asn Asp Lys Thr Ile Asp Glu Glu Leu Glu Arg Asp
          125                                130                                135                                140

aag agg gtc act tgg att gtg gag ttc ttt gcc aan tgg tct aat gac      577
Lys Arg Val Thr Trp Ile Val Glu Phe Phe Ala Xaa Trp Ser Asn Asp
          145                                150                                155

tgc caa tca ttt gcc cct atc tat gct gac ctc tcc ctt aaa tac aac      625
Cys Gln Ser Phe Ala Pro Ile Tyr Ala Asp Leu Ser Leu Lys Tyr Asn
          160                                165                                170

tgt aca ggg cta aat ttt ggg aag gtg gat gtt gga cgc tat act gat      673
Cys Thr Gly Leu Asn Phe Gly Lys Val Asp Val Gly Arg Tyr Thr Asp
          175                                180                                185

gtt agt acg cgg tac aaa gtg agc aca tca ccc ctc acc aag caa ctc      721
Val Ser Thr Arg Tyr Lys Val Ser Thr Ser Pro Leu Thr Lys Gln Leu
          190                                195                                200

cct acc ctg atc ctg ttc caa ggt ggc aag gag gca atg cgg cgg cca      769
Pro Thr Leu Ile Leu Phe Gln Gly Gly Lys Glu Ala Met Arg Arg Pro
          205                                210                                215                                220

cag at
Gln
          774

```

&lt;210&gt; 287

<211> 614  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 155..613

<221> sig\_peptide  
 <222> 155..205  
 <223> Von Heijne matrix  
 score 5.80000019073486  
 seq LWLKLLAFGFAFL/DT

<400> 287  
 aaaaaccgaa tctgacatca tcacctagca gttcatgcag ctagcaagtg gtttggttctt 60  
 agggtaacag aggaggaaat tggttcctcgt ctgataagac aacagtggag aaaggacgca 120  
 tgctgtttct tagggacacg gctgacttcc agat atg acc atg tat ttg tgg ctt 175  
 Met Thr Met Tyr Leu Trp Leu  
 -15  
 aaa ctc ttg gca ttt ggc ttt gcc ttt ctg gac aca gaa gta ttt gtg 223  
 Lys Leu Leu Ala Phe Gly Phe Ala Phe Leu Asp Thr Glu Val Phe Val  
 -10 -5 1 5  
 aca ggg caa agc cca aca cct tcc ccc act gga ttg act aca gca aag 271  
 Thr Gly Gln Ser Pro Thr Pro Ser Pro Thr Gly Leu Thr Thr Ala Lys  
 10 15 20  
 atg ccc agt gtt cca ctt tca agt gac ccc tta cct act cac acc act 319  
 Met Pro Ser Val Pro Leu Ser Ser Asp Pro Leu Pro Thr His Thr Thr  
 25 30 35  
 gca ttc tca ccc gca agc acc ttt gaa aga gaa aat gac ttc tca gag 367  
 Ala Phe Ser Pro Ala Ser Thr Phe Glu Arg Glu Asn Asp Phe Ser Glu  
 40 45 50  
 acc aca act tct ctt agt cca gac aat act tcc acc caa gta tcc ccg 415  
 Thr Thr Thr Ser Leu Ser Pro Asp Asn Thr Ser Thr Gln Val Ser Pro  
 55 60 65 70  
 gac tct ttg gat aat gct agt gct ttt ark acc aca ggt gtt tca tca 463  
 Asp Ser Leu Asp Asn Ala Ser Ala Phe Xaa Thr Thr Gly Val Ser Ser  
 75 80 85  
 gta cag acg cct cas ctt ccc acg cac gca gac tcg cag acg ccc tct 511  
 Val Gln Thr Pro Xaa Leu Pro Thr His Ala Asp Ser Gln Thr Pro Ser  
 90 95 100  
 gct gga act gac acg cag aca ttc agc ggc tcc gcg sca atg caa aac 559  
 Ala Gly Thr Asp Thr Gln Thr Phe Ser Gly Ser Ala Xaa Met Gln Asn  
 105 110 115  
 tca acc cta ccc cag gca gca atg cta tct cag atg tcc cag gag aga 607  
 Ser Thr Leu Pro Gln Ala Ala Met Leu Ser Gln Met Ser Gln Glu Arg  
 120 125 130  
 gga gta c 614  
 Gly Val  
 135

<210> 288  
 <211> 251  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 124..249

<221> sig\_peptide  
 <222> 124..174  
 <223> Von Heijne matrix

score 5.80000019073486

seq LWLKLLAFGFAFL/DT

&lt;400&gt; 288

```

atgtttttttt actttttacat ttttgattcg tttttacaga gaaaaacttc tacagagata      60
acaattatttt tgcttttcag aaggacgcat gctgtttctt agggacacgg ctgacttcca      120
gat atg acc atg tat ttg tgg ctt aaa ctc ttg gca ttt ggc ttt gcc      168
    Met Thr Met Tyr Leu Trp Leu Lys Leu Leu Ala Phe Gly Phe Ala
          -15                -10                -5
ttt ctg gac aca gaa gta ttt gtg aca ggg caa agc cca aca cct tcc      216
Phe Leu Asp Thr Glu Val Phe Val Thr Gly Gln Ser Pro Thr Pro Ser
          1                5                10
ccc act ggt gtt tca tca gta cag acg ccc cag gg      251
Pro Thr Gly Val Ser Ser Val Gln Thr Pro Gln
    15                20                25

```

&lt;210&gt; 289

&lt;211&gt; 416

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 155..415

&lt;221&gt; sig\_peptide

&lt;222&gt; 155..205

&lt;223&gt; Von Heijne matrix

score 5.80000019073486

seq LWLKLLAFGFAFL/DT

&lt;400&gt; 289

```

aaaaaccgaa tctgacatca tcacctagca gttcatgcag ctagcaagtg gtttgttctt      60
agggtaacag aggaggaaat tgttcctcgt ctgataagac aacagtggag aaaggacgca      120
tgctgtttct tagggacacg gctgacttcc agat atg acc atg tat ttg tgg ctt      175
                                Met Thr Met Tyr Leu Trp Leu
                                      -15
aaa ctc ttg gca ttt ggc ttt gcc ttt ctg gac aca gaa gta ttt gtg      223
Lys Leu Leu Ala Phe Gly Phe Ala Phe Leu Asp Thr Glu Val Phe Val
-10                -5                1                5
aca ggg caa agc cca aca cct tcc ccc act ggt gtt tca tca gta cag      271
Thr Gly Gln Ser Pro Thr Pro Ser Pro Thr Gly Val Ser Ser Val Gln
          10                15                20
acg cct cac ctt ccc acg cac gca gac tcg cag acg ccc tct gct gga      319
Thr Pro His Leu Pro Thr His Ala Asp Ser Gln Thr Pro Ser Ala Gly
          25                30                35
act gac acg cag aca ttc agc ggc tcc gcg sca atg caa aac tca acc      367
Thr Asp Thr Gln Thr Phe Ser Gly Ser Ala Xaa Met Gln Asn Ser Thr
          40                45                50
cta ccc cag gca gca atg cta tct cag atg tcc cag gag aga gga gta c      416
Leu Pro Gln Ala Ala Met Leu Ser Gln Met Ser Gln Glu Arg Gly Val
    55                60                65                70

```

&lt;210&gt; 290

&lt;211&gt; 309

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 183..308

&lt;221&gt; sig\_peptide

&lt;222&gt; 183..290

&lt;223&gt; Von Heijne matrix

score 5.80000019073486

seq LFLLSGTIWIAIC/KP

&lt;400&gt; 290

```

gaggcctttt ggtcacttag aagaggtgcc aaagatcaag gagaggaaag tggtagggcta    60
caaagtgtaaa ttctgtgtgg aagtgcaccc aacgctccga gccatctgca atcacctccg    120
wwagcacgctc cagtatggca atgtcccagc tgtgtcagct gctgtgaagg ggctgcgttc    180
tc atg aga gga gcc acc tgg ccc tgg cca tgt tta ccc gcg agg aca    227
  Met Arg Gly Ala Thr Trp Pro Trp Pro Cys Leu Pro Ala Arg Thr
      -35          -30          -25
agt aca gct gcc agt att gct cgt ttg ttt ctg ctt tca ggc aca att    275
Ser Thr Ala Ala Ser Ile Ala Arg Leu Phe Leu Leu Ser Gly Thr Ile
      -20          -15          -10
tgg atc gcc ata tgc aaa ccc acc acg aac ggg g    309
Trp Ile Ala Ile Cys Lys Pro Thr Thr Asn Gly
      -5          1          5

```

&lt;210&gt; 291

&lt;211&gt; 359

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 142..357

&lt;221&gt; sig\_peptide

&lt;222&gt; 142..333

&lt;223&gt; Von Heijne matrix

score 5.80000019073486

seq SALLRSLLLPXLX/QI

&lt;221&gt; misc\_feature

&lt;222&gt; 282..283

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 291

```

caagtccaca gcgctgggtg cagtacctcc ggcttcttca ggagtccatc tggcctggtg    60
gagttttgcc taagtttcca cggcccshta aggacccaag agcagaaact ggctgctgag    120
aaacaggcctt tgcagagcct g atg gga gtc ctc cca gat ctc gta gta gaa    171
              Met Gly Val Leu Pro Asp Leu Val Val Glu
                  -60          -55
att ttt ggg gtg aac aaa tgc cgg ctg agc tgg ggt cta gtc ctg gag    219
Ile Phe Gly Val Asn Lys Cys Arg Leu Ser Trp Gly Leu Val Leu Glu
      -50          -45          -40
tca cta caa caa ccc ctc atc aac agg cat ttg att tac tgc ctt ggg    267
Ser Leu Gln Gln Pro Leu Ile Asn Arg His Leu Ile Tyr Cys Leu Gly
      -35          -30          -25
gac atc atc ctg grn ntc ttg gat ctc agt gct ctg ttg agg agt ctg    315
Asp Ile Ile Leu Xaa Xaa Leu Asp Leu Ser Ala Leu Leu Arg Ser Leu
      -20          -15          -10
ctg cta cca sct ctg sct cag ata ccc cag gca act cta aga gg    359
Leu Leu Pro Xaa Leu Xaa Gln Ile Pro Gln Ala Thr Leu Arg
      -5          1          5

```

&lt;210&gt; 292

&lt;211&gt; 254

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 129..254

&lt;221&gt; sig\_peptide

&lt;222&gt; 129..173

&lt;223&gt; Von Heijne matrix

score 5.80000019073486

seq ALGFVLLAPRGWG/SL

&lt;400&gt; 292

```

gttttttcagc tcgccattca ctctcgctgtg aagatggcgt cgggcagcgg gacaaaaaac      60
ttggacttttc gccgaaagtg ggatgtggga aggtgggcag ggaccagatc aaaggagaca      120
gccaggag atg aca gca ctg ggg ttt gtt ctg tta gct cca cgt ggc tgg      170
      Met Thr Ala Leu Gly Phe Val Leu Leu Ala Pro Arg Gly Trp
      -15              -10              -5
ggg agc ctc aca gtc atg gtg gaa ggc aag gaa gag caa gtc acg tct      218
Gly Ser Leu Thr Val Met Val Glu Gly Lys Glu Glu Gln Val Thr Ser
      1              5              10              15
tac acg gat ggc agc agg caa aga gac agc aat ttt      254
Tyr Thr Asp Gly Ser Arg Gln Arg Asp Ser Asn Phe
      20              25

```

&lt;210&gt; 293

&lt;211&gt; 414

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 221..412

&lt;221&gt; sig\_peptide

&lt;222&gt; 221..337

&lt;223&gt; Von Heijne matrix

score 5.80000019073486

seq LISFLHTLQVVCS/VI

&lt;400&gt; 293

```

gtcagagcac atccggtgtt agaagcgctg gtaggccttg gagaggcggg ttaggaagag      60
tggagactgc tgcacggact ctggaaccat gaacatatat gatcgaaaga tcaactttga      120
tgcgctttta aaatttttctc atataacccc gtcaacgcag cagsrcctga agaagatttc      180
attactgtct tcagaaaact catgatgata ctggccatga atg aaa agg ata aga      235
                        Met Lys Arg Ile Arg
                        -35
aga aag aga aga aat gaa gtg acc atc cag cct ttc cca att aga ctt      283
Arg Lys Arg Arg Asn Glu Val Thr Ile Gln Pro Phe Pro Ile Arg Leu
      -30              -25              -20
cct ctc ctt cca ccc ctc att tcc ttt ttg cac aca tta cag gtg gtg      331
Pro Leu Leu Pro Pro Leu Ile Ser Phe Leu His Thr Leu Gln Val Val
      -15              -10              -5
tgt tct gtg ata atg aaa agc atc aga aaa gct ttt gta ctt tgt ggt      379
Cys Ser Val Ile Met Lys Ser Ile Arg Lys Ala Phe Val Leu Cys Gly
      1              5              10
ttc ctc tat ttt gaa ttt ttt gat caa aaa ctg at      414
Phe Leu Tyr Phe Glu Phe Phe Asp Gln Lys Leu
      15              20              25

```

&lt;210&gt; 294

&lt;211&gt; 334

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 156..332

&lt;221&gt; sig\_peptide

&lt;222&gt; 156..221

&lt;223&gt; Von Heijne matrix

score 5.80000019073486

seq XVXLFXXVXXA/AL

&lt;400&gt; 294

catgtgtttc tcatctttca accaccaaac ttacataaaa ttcattccacc tteccattct 60

tcctattata gtacaggaaa ggtccttcct gtcaaaggca aaatcactta tgattgtgtc 120

cccatctctc ttgccttttc aaggactttg agcct atg ctg cca ctg ctt cat 173

Met Leu Pro Leu Leu His

-20

tgt ttt ttt ttk gtt kgt ttg ttt kgt ttk gtt ttk gtt twa ama gca 221

Cys Phe Phe Xaa Val Xaa Leu Phe Xaa Xaa Val Xaa Val Xaa Xaa Ala

-15

-10

-5

gct tta ttg aga tat aat yca agt ata cag kgt ggc cgg gca cag kgg 269

Ala Leu Leu Arg Tyr Asn Xaa Ser Ile Gln Xaa Gly Arg Ala Gln Xaa

1 5 10 15

ctc ama cct gwa atc cca gma ctt tgg gag act aag gma ggc aga tta 317

Leu Xaa Pro Xaa Ile Pro Xaa Leu Trp Glu Thr Lys Xaa Gly Arg Leu

20

25

30

ctt gag cct agg aat tt

Leu Glu Pro Arg Asn

35

334

&lt;210&gt; 295

&lt;211&gt; 375

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 284..373

&lt;221&gt; sig\_peptide

&lt;222&gt; 284..346

&lt;223&gt; Von Heijne matrix

score 5.80000019073486

seq EMCLSQLFLQVEA/AY

&lt;400&gt; 295

ataacgctac ctgccgagca agccgagtgg agggaggcaa aggccaggct gaggatcagg 60

gtggcccggtg tggcagcggg gaggcgctgc atgctggagg ctgtgctgag tgcccggtgc 120

aggtgagccg gtctctgcga gttgtgccga gtgcctgctg cagtctcatt tccagttcct 180

ccatgacgtg gcaagtgaag acaggaatga aaggratgta aagcagcttt tctctgaaga 240

gaagaagaga gagagacaca gccaaagaccg aggctgggcc aag atg gtg tct gtg 295

Met Val Ser Val

-20

ttt cga agc gag gag atg tgt ttg tca caa ctg ttt ctc cag gtg gaa 343

Phe Arg Ser Glu Glu Met Cys Leu Ser Gln Leu Phe Leu Gln Val Glu

-15

-10

-5

gct gca tat tgc tgt gtg gct gag ctc gga ga

Ala Ala Tyr Cys Cys Val Ala Glu Leu Gly

1

5

375

&lt;210&gt; 296

&lt;211&gt; 226

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

<220>  
 <221> CDS  
 <222> 104..226

<221> sig\_peptide  
 <222> 104..187  
 <223> Von Heijne matrix  
 score 5.80000019073486  
 seq WWSVASLLSDVAA/WW

<400> 296  
 tggattgaaa taaattccta gctccacggt caggtcagta ggctgccatg atgaaatttg 60  
 aagaagagtc tgttatgatg tgtaatacca atttctggag ggc atg gct gct ctc 115  
 Met Ala Ala Leu  
 -25  
 cga agt act cta aca tgg aca gaa gtc gtg ggc tgg tgg agt gtt gcg 163  
 Arg Ser Thr Leu Thr Trp Thr Glu Val Val Gly Trp Trp Ser Val Ala  
 -20 -15 -10  
 tcg ctg ctt agt gat gtg gca gca tgg tgg cca ccg cac tcc acc tca 211  
 Ser Leu Leu Ser Asp Val Ala Ala Trp Trp Pro Pro His Ser Thr Ser  
 -5 1 5  
 aca cgg gga ggg gta 226  
 Thr Arg Gly Gly Val  
 10

<210> 297  
 <211> 232  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 90..230  
 <221> sig\_peptide  
 <222> 90..221  
 <223> Von Heijne matrix  
 score 5.80000019073486  
 seq LVCVFTCSLLAFF/SP

<400> 297  
 ctgaactttt tatttttcta tttttataac caggagaaaag taaacacata cacacacatg 60  
 gatggagaga gggacagagg gatggacgg atg aat gca ttr gta gat ggg aaa 113  
 Met Asn Ala Leu Val Asp Gly Lys  
 -40  
 cgg ctt asa krg tgc ata cgc tat ttc gat tct atc tca cta tat tct 161  
 Arg Leu Xaa Xaa Cys Ile Arg Tyr Phe Asp Ser Ile Ser Leu Tyr Ser  
 -35 -30 -25  
 aag gca agt tta agt tgt tgt tta gtg tgt gtg ttt act tgt tca ttg 209  
 Lys Ala Ser Leu Ser Cys Cys Leu Val Cys Val Phe Thr Cys Ser Leu  
 -20 -15 -10 -5  
 cta gct ttc ttc agc cca tgc ac 232  
 Leu Ala Phe Phe Ser Pro Cys  
 1

<210> 298  
 <211> 258  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS





168

<223> Von Heijne matrix  
score 5.80000019073486  
seq SFYFLASSSLSTS/AS

<221> misc\_feature  
<222> 16,286  
<223> n=a, g, c or t

<400> 300  
asatcgcgct gggganasgc cacgtcgcta tgagtgtgtt tcagtctacc tggattaaac 60  
gtttgtcttct cttcgctctac cttgattaa cgtgcacttc gcagtcctcg gttctccata 120  
cccgtgacct ggggatcgct acggacctta aaataccgcg aacascctt tcgtsccaag 180  
ctggagagca gtggc atg atc tcg gct cac tgc agc ttc tac ttc ctg gcc 231  
Met Ile Ser Ala His Cys Ser Phe Tyr Phe Leu Ala  
-15 -10  
tca agc agt ctt tcc acc tca gcs tct saa cgc act gga att aca gat 279  
Ser Ser Ser Leu Ser Thr Ser Ala Ser Xaa Arg Thr Gly Ile Thr Asp  
-5 1 5  
gtg agc n 286  
Val Ser  
10

<210> 301  
<211> 242  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 113..241

<221> sig\_peptide  
<222> 113..184  
<223> Von Heijne matrix  
score 5.69999980926514  
seq CLFVFLYTPCNC/FG

<400> 301  
tgcaatgggt ggtgttttat aatgctctct tccctaataca tgtaatacag gagattttcc 60  
ttttggaact cctgactgaa agcttcttag ttacacaca tggtcctcca gg atg aac 118  
Met Asn  
gct gaa aat aac ttt ttc ggt ttt gtt tgt ttg ttt gtt ttc ctc tat 166  
Ala Glu Asn Asn Phe Phe Gly Phe Val Cys Leu Phe Val Phe Leu Tyr  
-20 -15 -10  
aca acc cct tgc aat tgc ttt ggt tta gaa cac ctt tgg att cta agt 214  
Thr Thr Pro Cys Asn Cys Phe Gly Leu Glu His Leu Trp Ile Leu Ser  
-5 1 5 10  
ttc atg gtt gtt ctg gga gwy acc agg g 242  
Phe Met Val Val Leu Gly Xaa Thr Arg  
15

<210> 302  
<211> 136  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 42..134

<221> sig\_peptide  
<222> 42..110

<223> Von Heijne matrix  
score 5.69999980926514  
seq LPCCCHLLTCVSS/LR

<400> 302  
agtcacagtg acacagcctt ccaaccaggc cgccccctgg c atg acc atg gct gtg 56  
Met Thr Met Ala Val  
-20  
ggt gca gct gmy cam ctc ccc tgc tgc tgc cat ttr ctc acc tgc gtm 104  
Gly Ala Ala Xaa Leu Pro Cys Cys Cys His Leu Leu Thr Cys Val  
-15 -10 -5  
tcc agc ctt cgc amt gac att tac cca cat gg 136  
Ser Ser Leu Arg Xaa Asp Ile Tyr Pro His  
1 5

<210> 303  
<211> 175  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 73..174

<221> sig\_peptide  
<222> 73..147  
<223> Von Heijne matrix  
score 5.69999980926514  
seq SILLAALSRNISP/GQ

<400> 303  
aagagggaagc ggaakdgcct cagggtggcg gtagtgccaa aagcccaggc cgtccgcgca 60  
aaccgaggcg tc atg cgg aga aaa agg cga gaa aga aaa gag agg aag agc 111  
Met Arg Arg Lys Arg Arg Glu Arg Lys Glu Arg Lys Ser  
-25 -20 -15  
atc ctc ctg gcc gcc ctt tcg agg aac ata agt cct ggt cag aca tac 159  
Ile Leu Leu Ala Ala Leu Ser Arg Asn Ile Ser Pro Gly Gln Thr Tyr  
-10 -5 1  
cga aca tcc ccc gcg g 175  
Arg Thr Ser Pro Ala  
5

<210> 304  
<211> 493  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 402..491

<221> sig\_peptide  
<222> 402..470  
<223> Von Heijne matrix  
score 5.69999980926514  
seq LELLTSSDPPLA/SQ

<400> 304  
ttaggtgttc tgatagttaa gtggtagtat catggtctta atttttcctt gaagtggcctt 60  
ttgatttgca tttccttaat gactaattag gttgagcatc ttttcattgta cttactggcc 120  
ttctttggag aaataccttt tccaaatcca atgggttgtc tttttttatt gttgatctta 180  
agggttctta ggtgttcttg gtaccagttt cttgtgagat gtgtgacttg taaatacttt 240  
cttccattct ccatgttgtc tttttattct cttgatggta ttctttgaaa tacaaaartk 300

170

```

tttatatttg acaaagttca gtttatttat ttatttattg ccattcgtgc ttttggtttt 360
gataatccat ttttwtgtt tttattttta tttacttaga g atg ggg tct ccc tat 416
                                         Met Gly Ser Pro Tyr
                                         -20
gtt gcc cac gtt ggt ctt gaa ctc ttg acc tca agt gat cct ccc tcc 464
Val Ala His Val Gly Leu Glu Leu Leu Thr Ser Ser Asp Pro Pro Ser
          -15                      -10                      -5
ttg gcc tcc caa gtg ctg gga ata cat tm 493
Leu Ala Ser Gln Val Leu Gly Ile His
      1                      5

```

<210> 305  
 <211> 214  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 79..213

<221> sig\_peptide  
 <222> 79..135  
 <223> Von Heijne matrix  
 score 5.69999980926514  
 seq VCWLTLTLAHSLS/LT

```

<400> 305
cacacacgca ccaaatacac acagasaccc tggccctcac tcacgcacav tctctcacac 60
tcgtggacac acccccag atg cat ctt tac act cat gta tgc tgg ctc act 111
                                         Met His Leu Tyr Thr His Val Cys Trp Leu Thr
                                         -15
ctc aca ctg gca cac tca cac agc ttg acc cac acg cac aca ctc aca 159
Leu Thr Leu Ala His Ser His Ser Leu Thr His Thr His Thr Leu Thr
          -5                      1                      5
ccc agt cac aca cgt aca cac tca cat acg tgt gct tgc cta cac gca 207
Pro Ser His Thr Arg Thr His Ser His Thr Cys Ala Cys Leu His Ala
      10                      15                      20
cac aag g 214
His Lys
25

```

<210> 306  
 <211> 458  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 306..458

<221> sig\_peptide  
 <222> 306..350  
 <223> Von Heijne matrix  
 score 5.69999980926514  
 seq LSLTFYHFPLCWG/HQ

<221> misc\_feature  
 <222> 286,448  
 <223> n=a, g, c or t

```

<400> 306
atcagagagc gccggaagcg gtccgagaat gaagcagtgat gatctaccat gcattgtctc 60

```

171

```

agaaagaggc gaatgactcc gatgtccagg tcagttcttg gcagggagtc caggagcaac 120
agaggtgatg gcaaagatgg ctccagtagct tctgagcccc cagcactgat tgagatgtcc 180
tttcccacat catactcctc atttttcttg cagacatcta aggctggatc aaagtctgta 240
gttctcatta cctgttccca cgtgccagcc tccttttctg ttgtgnmmaa gtcaagtttg 300
gtaaa atg agg ctt tcc tta acc ttt tat cat ttc cca ctg tgt tgg gga 350
      Met Arg Leu Ser Leu Thr Phe Tyr His Phe Pro Leu Cys Trp Gly
      -15                -10                -5
cac cag gct gtg ccc acg tgg tgg saa rgc atc att caa cct tgt cac 398
His Gln Ala Val Pro Thr Trp Trp Xaa Xaa Ile Ile Gln Pro Cys His
1          5          10          15
tgt gcc ctc tgc act tct gca gaa ggt gtg caa tca cat atc ata agt 446
Cys Ala Leu Cys Thr Ser Ala Glu Gly Val Gln Ser His Ile Ile Ser
      20          25          30
gna att tac aga 458
Xaa Ile Tyr Arg
      35

```

<210> 307  
 <211> 328  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 87..326

<221> sig\_peptide  
 <222> 87..128  
 <223> Von Heijne matrix  
       score 5.69999980926514  
       seq NVLIIVFVAFAG/FL

```

<400> 307
tatcttctct ccagtctaaa gcctcactga acaaactgtc cttgactgtc agtgctcagg 60
gaactgctct gccacccttc tcctca atg aat gtg tta atc att gtt ttt gtt 113
                        Met Asn Val Leu Ile Ile Val Phe Val
                        -10
gca ttt gct ttt ggg ttc ytg gtc atg aag tct ttg ctt aag cca atg 161
Ala Phe Ala Phe Gly Phe Leu Val Met Lys Ser Leu Leu Lys Pro Met
-5          1          5          10
tcg aga agg gtt ttt ctg atg tta tct tct agg att ttt atg gtt tca 209
Ser Arg Arg Val Phe Leu Met Leu Ser Ser Arg Ile Phe Met Val Ser
      15          20          25
ggg ctt aga ttt aag tcc ttg atc cat ctt gag ttg att ttt gta tat 257
Gly Leu Arg Phe Lys Ser Leu Ile His Leu Glu Leu Ile Phe Val Tyr
      30          35          40
aag ttg aga gat gag gat cca gtt tca ttc ttc tac atg tgg ctt gcc 305
Lys Leu Arg Asp Glu Asp Pro Val Ser Phe Phe Tyr Met Trp Leu Ala
      45          50          55
aat tat ccc agc acc att tgt tg 328
Asn Tyr Pro Ser Thr Ile Cys
60          65

```

<210> 308  
 <211> 380  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 33..380

<221> sig\_peptide

172

&lt;222&gt; 33..92

&lt;223&gt; Von Heijne matrix

score 5.69999980926514

seq LAWALPSLLRLGA/AQ

&lt;221&gt; misc\_feature

&lt;222&gt; 326

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 308

```

agcgggtctcc cggccctgc cgccctgcc ct atg tcc cgc cgc tct atg ctg      53
                               Met Ser Arg Arg Ser Met Leu
                               -20                               -15
ctt gcc tgg gct ctc ccc agc ctc ctt cga ctc gga gcg gct cag gag      101
Leu Ala Trp Ala Leu Pro Ser Leu Leu Arg Leu Gly Ala Ala Gln Glu
                               -10                               -5                               1
aca gaa gac ccg gcc tgc tgc agc ccc ata gtg ccc cgg aac gag tgg      149
Thr Glu Asp Pro Ala Cys Cys Ser Pro Ile Val Pro Arg Asn Glu Trp
                               5                               10                               15
aag gcc ctg gca tca gag tgc gcc cag cac ctg agc ctg ccc tta cgc      197
Lys Ala Leu Ala Ser Glu Cys Ala Gln His Leu Ser Leu Pro Leu Arg
20                               25                               30                               35
tat gtg gtg gta tgc cac acg gcg ggc agc agc tgc aac acc scc gcc      245
Tyr Val Val Val Ser His Thr Ala Gly Ser Ser Cys Asn Thr Xaa Ala
                               40                               45                               50
tcg tgc cag cag cag gcc cgg aat gtg cag cac tac cac atg aag aca      293
Ser Cys Gln Gln Gln Ala Arg Asn Val Gln His Tyr His Met Lys Thr
                               55                               60                               65
ctg ggc tgg tgc gac gtg ggc tac aac tkc ctn gat tgg aga aga cgg      341
Leu Gly Trp Cys Asp Val Gly Tyr Asn Xaa Leu Asp Trp Arg Arg Arg
70                               75                               80
gct cgt ata cra ggg ccg tgg mtg gaa ctt cac ggg tsc      380
Ala Arg Ile Xaa Gly Pro Trp Xaa Glu Leu His Gly Xaa
85                               90                               95

```

&lt;210&gt; 309

&lt;211&gt; 284

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 228..284

&lt;221&gt; sig\_peptide

&lt;222&gt; 228..269

&lt;223&gt; Von Heijne matrix

score 5.69999980926514

seq VFLFLMISVFAGC/QI

&lt;400&gt; 309

```

aaaagaagaa agctgaatca tactcgatga ttattgatca tttgtataca gctcaagccc      60
tcaagtagcc tgctgtaata ttactagtt acaaagaaaa gattcgtttt gtcacagtta      120
catgaaaggt gcttatatgt gcaaatatgg agacaaagtt catcttaaaa gattaaaatg      180
agaatctcct aaatgaagca tttggaatat tgattagtat accagaa atg gtt ttt      236
                               Met Val Phe
ctt ttt ctt atg atc agc gtt ttt gcc ggt tgt caa atc cct tcc ggg      284
Leu Phe Leu Met Ile Ser Val Phe Ala Gly Cys Gln Ile Pro Ser Gly
-10                               -5                               1                               5

```

&lt;210&gt; 310

&lt;211&gt; 357

<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 243..356

<221> sig\_peptide  
<222> 243..305  
<223> Von Heijne matrix  
score 5.59999990463257  
seq AGLELLASSNSSA/LP

```

<400> 310
ttgagatcac ctgaggcaac atagtgaac cctgtatcta gaataaatta gagaaagaaa      60
aatagtctgg gcatgatggt gtgcacctat agtctccagc tabtcasgag cctgaggcag      120
gaggwtcact tgagctkagg agttcaagga tgcagtsacc tgtgattgca ccaactgcatt      180
ccagcttgga caacagagtg agaccctgtc ttaaaattta aattttktgt yttwtggtag      240
ag atg ggg tct cgc cct gtt tcc gak gct ggt ctc gaa ctc ctg gcc      287
    Met Gly Ser Arg Pro Val Ser Xaa Ala Gly Leu Leu Leu Ala
      -20          -15          -10
tcg agc aat tct tct gcc ttg ccc ttc caa tgt tct ggg att aca ggc      335
Ser Ser Asn Ser Ser Ala Leu Pro Phe Gln Cys Ser Gly Ile Thr Gly
      -5          1          5          10
atg agc crc cac acc cta gcg g      357
Met Ser Xaa His Thr Leu Ala
      15

```

<210> 311  
<211> 470  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 413..469

<221> sig\_peptide  
<222> 413..451  
<223> Von Heijne matrix  
score 5.59999990463257  
seq MLCHLSLVFLGXG/QF

<221> misc\_feature  
<222> 30  
<223> n=a, g, c or t

```

<400> 311
ccgttaacgg gattctggaa tttgttaggn taattgcttt tcaatatcaa gagatctggc      60
aatcaaattt aataatatca agcttgcttg gtgagcatgg atttataaga tagaatgggt      120
tgtgggggrg artatagtkc cgaaaaagrk tattgtttcc cataatgcct ggtattgtat      180
taagtacttt gcatacagta gggcatttca ttgtcccagt gatcctcctg caaagtaggt      240
acaattatct tcaatttaca aatgaggaaa ccaagctctc ttcaagctga taagatgctg      300
aactgagatt tgaaccaagt ccctctgccc ctaagagccc ctaccctag ctgctactat      360
atgctgtacc catctaagct ttgtgaaata rccttggtcc actgcagaga ag atg ttg      418
                                Met Leu
tgt cac cta tct cta gta ttt ctt ggc ktt ggg cag ttc tgg agt caa      466
Cys His Leu Ser Leu Val Phe Leu Gly Xaa Gly Gln Phe Trp Ser Gln
      -10          -5          1          5
aat g      470
Asn

```

<210> 312  
 <211> 187  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 98..187

<221> sig\_peptide  
 <222> 98..148  
 <223> Von Heijne matrix  
 score 5.59999990463257  
 seq FMCLFAICISSNA/KC

<400> 312  
 aagtttggttt ttgttggtgg tggatgtagt ttcttatatt ctattccata aagtatgaaa 60  
 tggaggctcc ttgtgatttt taatttgcac ttctgta atg act aat ctt ttc atg 115  
 Met Thr Asn Leu Phe Met  
 -15  
 tgc ttg ttt gcc atc tgt ata tct tct aat gcg aag tgt ctg ttt agt 163  
 Cys Leu Phe Ala Ile Cys Ile Ser Ser Asn Ala Lys Cys Leu Phe Ser  
 -10 -5 1 5  
 ctt ttt cct ttt ttt att gag ggg 187  
 Leu Phe Pro Phe Phe Ile Glu Gly  
 10

<210> 313  
 <211> 237  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 93..236

<221> sig\_peptide  
 <222> 93..173  
 <223> Von Heijne matrix  
 score 5.59999990463257  
 seq CVLFTLLVSTRSG/RS

<221> misc\_feature  
 <222> 111  
 <223> n=a, g, c or t

<400> 313  
 ttgcttagga ttttctaaaa gattacataa aatactgttg aaaagatgat tgcatacaaa 60  
 acataatctg ttcattgtta aacgtatacg aa atg ttg gga tac atc tgg naa 113  
 Met Leu Gly Tyr Ile Trp Xaa  
 -25  
 caa gac aaa gtc ttt gct aat tgt gtt cta ttt acg ctc tta gtg tct 161  
 Gln Asp Lys Val Phe Ala Asn Cys Val Leu Phe Thr Leu Leu Val Ser  
 -20 -15 -10 -5  
 aca aga tcc ggg aga tcg cgs gcg ggt tgt gcc tgg agg tgg agg gga 209  
 Thr Arg Ser Gly Arg Ser Arg Ala Gly Cys Ala Trp Arg Trp Arg Gly  
 1 5 10  
 aga tgg tca gta gga cag aag ggc hgg g 237  
 Arg Trp Ser Val Gly Gln Lys Gly Xaa  
 15 20

<210> 314



175

<211> 356  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 272..355

<221> sig\_peptide  
 <222> 272..316  
 <223> Von Heijne matrix  
 score 5.59999990463257  
 seq LILSLQVCRPATL/DQ

<221> misc\_feature  
 <222> 275..276  
 <223> n=a, g, c or t

<400> 314  
 ggatttgctt tctttttctc caaaagggga ggaaattgaa actgagtggc ccacgatggg 60  
 aagaggggaa agcccagggg tacaggaggc ctctgggtga aggcagaggc taacatgggg 120  
 ttcgagcgga ccttggccgt tggcctgacc atctttgtgc tgtctgtcgt cactatcatc 180  
 atctgcttca cctgctcctg ctgctgcctt tacaagacgt gccgccgacc acgtccggtt 240  
 gtcaccacca ccacatccac cactgtggtg c atg nnc ctt atc ctc agc ctc 292  
 Met Xaa Leu Ile Leu Ser Leu  
 -15 -10  
 caa gtg tgc cgc cca gct acc ctg gac caa gct acc agg gct acc aca 340  
 Gln Val Cys Arg Pro Ala Thr Leu Asp Gln Ala Thr Arg Ala Thr Thr  
 -5 1 5  
 cca tgc cgc cta cgg g 356  
 Pro Cys Arg Leu Arg  
 10

<210> 315  
 <211> 162  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 40..162

<221> sig\_peptide  
 <222> 40..150  
 <223> Von Heijne matrix  
 score 5.59999990463257  
 seq VLMLSLPLPPTPQ/QA

<400> 315  
 tacatgtgca gaatgtgcag atttgtcaca taggtgtgt atg tgc cac agg cgt 54  
 Met Cys His Arg Arg  
 -35  
 tgg ctg cac cta tca acc cgt cat cta ggt ttt aag ccc cgc atc cat 102  
 Trp Leu His Leu Ser Thr Arg His Leu Gly Phe Lys Pro Arg Ile His  
 -30 -25 -20  
 tac gta ttt gtc tta atg ctg tcc ctc ccc ttg ccc ccc acc ccc caa 150  
 Tyr Val Phe Val Leu Met Leu Ser Leu Pro Leu Pro Pro Thr Pro Gln  
 -15 -10 -5  
 cag gcc ctc ggg 162  
 Gln Ala Leu Gly  
 1

176

<210> 316  
 <211> 404  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 297..404

<221> sig\_peptide  
 <222> 297..353  
 <223> Von Heijne matrix  
 score 5.59999990463257  
 seq FVIFPAALLLCWG/GL

<400> 316  
 taagctgaaa aagaatataa aaattaaaga gaaattgaaa atctaagtct tgcagtgaga 60  
 atgaccagaa atcgtttccc tctctggggg gttcctgttt aatatgaaag tcctcttaac 120  
 aagcgtggac agaggaagtt ttaggtttga tttgaacttc atgtacatga catatttcat 180  
 ttttttttct tccctcacia atttcaaccc aggccacttg tttgcagaga ctgccaaacc 240  
 ttccattgct gcttccaaga tactcctgga atctgagatt acctttttatc ctcttg atg 299  
 Met  
 gac cat gtt gtt att ttt gtc att ttc cct gca gct ctt ctg ctt tgc 347  
 Asp His Val Val Ile Phe Val Ile Phe Pro Ala Ala Leu Leu Leu Cys  
 -15 -10 -5  
 tgg gga gga ctc atc ccc cta tgc atc atc tac ccc ccg ata gct gac 395  
 Trp Gly Gly Leu Ile Pro Leu Cys Ile Ile Tyr Pro Pro Ile Ala Asp  
 1 5 10  
 aca gtt ggg 404  
 Thr Val Gly  
 15

<210> 317  
 <211> 450  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 359..448

<221> sig\_peptide  
 <222> 359..433  
 <223> Von Heijne matrix  
 score 5.59999990463257  
 seq LIIILXFDIYSLA/FI

<221> misc\_feature  
 <222> 323,410  
 <223> n=a, g, c or t

<400> 317  
 tatgtctttt gaatttgtga tgtacatatt aacagtagat taagttgaaa taataaaatc 60  
 tgtattgttt atgatttatt agttatatga tgagtagaat atagtctatt gtggscmagt 120  
 gtgtatatat aacataaaca atacattaac ccaattttgt gtgaaaatta ttttgggacc 180  
 tagtagcttt cttgggtcaca acctttcaaa caaacaattt tttttttaat taattttttc 240  
 ccttaataaa gaaaacaatt cctcaatgtg taatagcaaa taccttttaa caggatcatat 300  
 atcatcaatg ctttctttga aancgtactg atgctttaca gatgctttac gagtaaag 358  
 atg ctt aca aat ctt ttc ttt caa gta gct cat cct ctg atc att att 406  
 Met Leu Thr Asn Leu Phe Phe Gln Val Ala His Pro Leu Ile Ile Ile  
 -25 -20 -15 -10  
 ctg ntg ttt gat atc tac tcc cta gca ttt atc cat gac gtg gg 450

Leu Xaa Phe Asp Ile Tyr Ser Leu Ala Phe Ile His Asp Val  
-5 1 5

```
<221> sig_peptide
<222> 313..354
<223> Von Heijne matrix
      score 5.59999990463257
      seq LFGLRGMLPLTQQ/AP
```

```
<221> sig_peptide
<222> 20..82
<223> Von Heijne matrix
      score 5.59999990463257
      seq ACVYACVCASVSA/CV
```

[illegible]

Ser Val Cys Val  
55

<210> 320  
<211> 325  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 256..324  
  
<221> sig\_peptide  
<222> 256..318  
<223> Von Heijne matrix  
score 5.59999990463257  
seq LIANLVLPISIAA/LR

```

<400> 320
accacgcctc ctccaagtcc cagcgaaccc gcgtgcaacc tgtccctaaa aaagccaaag      60
cagtcactct ttacctccca ctttcctccc tcccagcctt tggcaaccac taatctactt      120
tccgtgtata tggatttacc tattcaggac atttcatatg tcctttggtg actggcttct      180
ttcactttgc acaatgtttt taagggtcat tcctgtcata gtgtgtgtca gtacgaaccc      240
ctccttaacc atcta atg gtt atc acc tct aat agt tat ctc ata gcc aat      291
                Met Val Ile Thr Ser Asn Ser Tyr Leu Ile Ala Asn
                  -20          -15          -10
ctt gtt tta ttt ata tct atc gcc gcc ctc cgg g      325
Leu Val Leu Phe Ile Ser Ile Ala Ala Leu Arg
                -5          1

```

<210> 321  
<211> 201  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 31..201

<221> sig\_peptide  
<222> 31..183  
<223> Von Heijne matrix  
score 5.5  
seq LSLHASLVTKAFS/IN

```

<400> 321
catcacaaga acccagagtg gaattctggg atg gaa gag ctg gac aga aag tgg      54
                Met Glu Glu Leu Asp Arg Lys Trp
                  -50          -45
aga gag aag gtc ctc cca gcg gca aag cta att aaa agg aga aac ctg      102
Arg Glu Lys Val Leu Pro Ala Ala Lys Leu Ile Lys Arg Arg Asn Leu
                -40          -35          -30
ttt tcc aca tgc act cct caa tat ggy aca cat gct gct ttc ttg tca      150
Phe Ser Thr Cys Thr Pro Gln Tyr Gly Thr His Ala Ala Phe Leu Ser
                -25          -20          -15
tta cat gcc tca ctt gtc acc aaa gca ttt tca atc aat tcc tgg gag      198
Leu His Ala Ser Leu Val Thr Lys Ala Phe Ser Ile Asn Ser Trp Glu
                -10          -5          1          5
tgg      201
Trp

```

<210> 322  
<211> 159

<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 77..157

<221> sig\_peptide  
<222> 77..151  
<223> Von Heijne matrix  
score 5.5  
seq PLLLCPLSSGSPC/PR

<400> 322  
aacaaggga cagaatgggc ccagggttcc ttctttcttcc ttccagttaa gagctcagag 60  
tggaagtggg ctgggg atg gtg tcg ggg gcc caa gct ccc agc tcc caa agg 112  
Met Val Ser Gly Ala Gln Ala Pro Ser Ser Gln Arg  
-25 -20 -15  
ccc ctg ctt cta tgc cct ttg agc tca ggt agc ccc tgc ccc cgg gg 159  
Pro Leu Leu Cys Pro Leu Ser Ser Gly Ser Pro Cys Pro Arg  
-10 -5 1

<210> 323  
<211> 420  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 325..420

<221> sig\_peptide  
<222> 325..405  
<223> Von Heijne matrix  
score 5.5  
seq SFLPSLLSSFLLS/LP

<221> misc\_feature  
<222> 117  
<223> n=a, g, c or t

<400> 323  
catgcaggat agtaatacgt tagaatcaaa aataagggtta tacttagaaa atattgattt 60  
gcctttttga ttttgcattg gtataatctg gctctgaaat cagtgcacacg aagtganctt 120  
cgaaacaagc ctgagcaata gaagtagatg tggaaataac ttcggtttct caaggcaaat 180  
actttgatag gaacaacaaa ccgttttagat atagaagatg tgatacatc ctttaaaaag 240  
aatttgacct tatgtcattg taggcacacc tcatatttca attattcata tagtttttct 300  
tgagcaattg ctggtttaag aata atg tca tgt ctt ttg cgt gct tat atc 351  
Met Ser Cys Leu Leu Arg Ala Tyr Ile  
-25 -20  
att tgg ata ttt cct tcc ttc ctt cct tcc ctc ctt tct tcc ttc ctt 399  
Ile Trp Ile Phe Pro Ser Phe Leu Pro Ser Leu Leu Ser Ser Phe Leu  
-15 -10 -5  
ctt tcc ctc ccc cct tcc ggg 420  
Leu Ser Leu Pro Pro Ser Gly  
1 5

<210> 324  
<211> 210  
<212> DNA  
<213> Homo sapiens

180

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 9..209

&lt;221&gt; sig\_peptide

&lt;222&gt; 9..116

&lt;223&gt; Von Heijne matrix

score 5.5

seq LHFVYCFLCCA EA/FL

&lt;400&gt; 324

```

ctccttat atg ttt cag tta ctg atc ctt tgt cag atg aat agt ttg aaa      50
      Met Phe Gln Leu Leu Ile Leu Cys Gln Met Asn Ser Leu Lys
      -35                    -30                    -25
ata ttt tct ccc att ctt gga tgg tct ctt cat ttt gtt tat tgt ttc      98
Ile Phe Ser Pro Ile Leu Gly Trp Ser Leu His Phe Val Tyr Cys Phe
      -20                    -15                    -10
ctt tgc tgt gca gaa gcc ttt tta ctt gat atg atc cca ttt atg caa      146
Leu Cys Cys Ala Glu Ala Phe Leu Leu Asp Met Ile Pro Phe Met Gln
      -5                    1                    5                    10
ttt tac ttt ggt tac ctg tgc ttg tgg ggt att act tta aaa atc ttt      194
Phe Tyr Phe Gly Tyr Leu Cys Leu Trp Gly Ile Thr Leu Lys Ile Phe
      15                    20                    25
gcc cag tcc aat tgg g
Ala Gln Ser Asn Trp
      30

```

&lt;210&gt; 325

&lt;211&gt; 192

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 31..192

&lt;221&gt; sig\_peptide

&lt;222&gt; 31..174

&lt;223&gt; Von Heijne matrix

score 5.5

seq VCLRLHVLSAVQT/ER

&lt;400&gt; 325

```

aggctgctgc agttggcgma tgaggcgacc atg gcc ttg ctg ggt aag cgc tgt      54
      Met Ala Leu Leu Gly Lys Arg Cys
      -45
gac gtc ccc acm aac ggc tgc gga ccc gac cgc wgg aam wac ggc gwy      102
Asp Val Pro Thr Asn Gly Cys Gly Pro Asp Arg Xaa Xaa Xaa Gly Xaa
-40                    -35                    -30                    -25
aac ccg caa ara cga gat cat cac cag cmt mgt gtc tgc ctt aga ctc      150
Asn Pro Gln Xaa Arg Asp His His Gln Xaa Xaa Val Cys Leu Arg Leu
      -20                    -15                    -10
cat gtg ctc agc gct gtc car act gaa cgc cga ggt gat ggg      192
His Val Leu Ser Ala Val Gln Thr Glu Arg Arg Gly Asp Gly
      -5                    1                    5

```

&lt;210&gt; 326

&lt;211&gt; 181

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

<222> 71..181

<221> sig\_peptide

<222> 71..166

<223> Von Heijne matrix  
score 5.5  
seq TLALLSSDSVATG/SV

<400> 326

```

aatttcgcgg cctagtgggg cgtagcgggcc tcttttgaaa gcctgagtta cgatgtattg      60
agcgcgtcgt atg cgg cca gca cta agg tcc ttc tgg cac tcc tct ggt      109
              Met Arg Pro Ala Leu Arg Ser Phe Trp His Ser Ser Gly
              -30              -25              -20
gga ccg ccc cca tcg gcc aca ctt gcc ctg ctc tcc agt gat tct gta      157
Gly Pro Pro Pro Ser Ala Thr Leu Ala Leu Leu Ser Ser Asp Ser Val
              -15              -10              -5
gct act ggc tcc gta gtc tcg cgg      181
Ala Thr Gly Ser Val Val Ser Arg
              1              5

```

<210> 327

<211> 185

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 39..185

<221> sig\_peptide

<222> 39..116

<223> Von Heijne matrix  
score 5.5  
seq LFSGWLVWWSRS/SQ

<221> misc\_feature

<222> 143,145,175

<223> n=a, g, c or t

<400> 327

```

caaagacgca ctacttagta cagagagggt ttgaatac atg ctc tgt gca tgc aag      56
              Met Leu Cys Ala Cys Lys
              -25
gca cgt ggg gtg atg ctg ctg ctg ttc tca ggg tgg ttg gtt tgg tgg      104
Ala Arg Gly Val Met Leu Leu Leu Phe Ser Gly Trp Leu Val Trp Trp
-20              -15              -10              -5
ggc agt agg tcc tca cag twc ctc aga atg cct gag agn tna gta agt      152
Gly Ser Arg Ser Ser Gln Xaa Leu Arg Met Pro Glu Xaa Xaa Val Ser
              1              5              10
ggg gag ggt cga agc gat cdv dng cca cat ggg      185
Gly Glu Gly Arg Ser Asp Xaa Xaa Pro His Gly
              15              20

```

<210> 328

<211> 210

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 57..209

&lt;221&gt; sig\_peptide

&lt;222&gt; 57..182

&lt;223&gt; Von Heijne matrix

score 5.5

seq SLILPTSPSPAHS/GS

&lt;400&gt; 328

gacttaggts yaggcgactg cccagacaat gactgggtccc gcataccgag cagagc atg 59  
Met

atc agc agc agt ctg agt gga aga gtg cct gtg atc tta ggg aac ctg 107  
Ile Ser Ser Ser Leu Ser Gly Arg Val Pro Val Ile Leu Gly Asn Leu

-40 -35 -30

atg ggc gtt gga gca gcg gtt cga cgc atg ggt ttc tct tta atc ctt 155  
Met Gly Val Gly Ala Ala Val Arg Arg Met Gly Phe Ser Leu Ile Leu

-25 -20 -15 -10

ccg act tcc cca agc cca gcg cac tca ggt tcc gct cca agt gcg gga 203  
Pro Thr Ser Pro Ser Pro Ala His Ser Gly Ser Ala Pro Ser Ala Gly

-5 1 5

ccc cgc g 210  
Pro Arg

&lt;210&gt; 329

&lt;211&gt; 318

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 149..316

&lt;221&gt; sig\_peptide

&lt;222&gt; 149..286

&lt;223&gt; Von Heijne matrix

score 5.5

seq ILLISTLFYSLLS/GS

&lt;400&gt; 329

acacttaacc catctgtttt ctctaattgca cgacagattc ctttcagaca ggacaactgt 60  
gatatttcag ttcttgattg taaatacctc ctaagcctga agcttctgtt actagccatt 120

gtgrgcttca gktctttcak yckgcaaa atg ggc ata ata car kct att ctt 172  
Met Gly Ile Ile Gln Xaa Ile Leu

-45 -40

gcc aca tca agg gat tgt tat tcc ttt aaa aaa aaa cca ata cca aag 220  
Ala Thr Ser Arg Asp Cys Tyr Ser Phe Lys Lys Lys Pro Ile Pro Lys

-35 -30 -25

aag cct aca atg ttg gcc tta gcc aaa att ctg ttg att tca acg ttg 268  
Lys Pro Thr Met Leu Ala Leu Ala Lys Ile Leu Leu Ile Ser Thr Leu

-20 -15 -10

ttt tat tca ctt cta tcg ggg agc cat gga aaa gra aat caa gac gtg 316  
Phe Tyr Ser Leu Leu Ser Gly Ser His Gly Lys Xaa Asn Gln Asp Val

-5 1 5 10

gg 318

&lt;210&gt; 330

&lt;211&gt; 223

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 135..221

&lt;221&gt; sig\_peptide



<222> 135..203  
 <223> Von Heijne matrix  
 score 5.5  
 seq LPFVCLLLRNVS/DL

<400> 330  
 aacagtgtgt gagagtcccc ttctctccac atcctcgcca gcatctgtta ttgcctgtct 60  
 ttttgatacg agccttttta acaggggtaa gatgatatct cattgtagtt ttgatttgca 120  
 ttctctgatg atca atg atg ttg agc acc ttt tca tat gcc tgt ttg cca 170  
 Met Met Leu Ser Thr Phe Ser Tyr Ala Cys Leu Pro  
 -20 -15  
 ttt gta tgt ctt ctt ttg aga aat gtc tat tca gat ctt ttg ccc aat 218  
 Phe Val Cys Leu Leu Leu Arg Asn Val Tyr Ser Asp Leu Leu Pro Asn  
 -10 -5 1 5  
 cgg gg 223  
 Arg

<210> 331  
 <211> 362  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 272..361

<221> sig\_peptide  
 <222> 272..343  
 <223> Von Heijne matrix  
 score 5.5  
 seq LIVVLVCISLVII/DD

<400> 331  
 aatggacacc taggttgctt ccatactctga gctattgtga ataatgctgc aatgaacatg 60  
 ggagtggaga catctcctaa gcatactgat ttcagttcct ttgggtatat acccagaagt 120  
 gggatcatgt ggtaatcttg tttttacttt tttgaggaac ctccatacca ttatccatga 180  
 tggctatagt aatttacatt cataccagca gtgcacaagg gtctcctttt ctgtatacac 240  
 ttgccaacac ttgttatctt tcattttttt g atg cta gcc att cta aca ggt 292  
 Met Leu Ala Ile Leu Thr Gly  
 -20  
 ggg agg tgg tat ctc ata gtg gtt tta gtt tgc att tcc ttg gtg att 340  
 Gly Arg Trp Tyr Leu Ile Val Leu Val Cys Ile Ser Leu Val Ile  
 -15 -10 -5  
 att gat gat gat gag cac ggg g 362  
 Ile Asp Asp Asp Glu His Gly  
 1 5

<210> 332  
 <211> 89  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 34..87

<221> sig\_peptide  
 <222> 34..75  
 <223> Von Heijne matrix  
 score 5.5  
 seq LLPLGLKVLGLQA/RG

<400> 332

184

cccagaccgg tcttgaactc ctggcctcaa ctg atg ctc ctg cct ctg ggt ctc 54  
Met Leu Leu Pro Leu Gly Leu  
-10

aaa gtg ctg gga tta cag gcg aga ggc acc acg ct 89  
Lys Val Leu Gly Leu Gln Ala Arg Gly Thr Thr  
-5 1

&lt;210&gt; 333

&lt;211&gt; 399

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 255..398

&lt;221&gt; sig\_peptide

&lt;222&gt; 255..338

&lt;223&gt; Von Heijne matrix

score 5.5

seq PTLVMMWLSPQMA/SS

&lt;400&gt; 333

ttcactgcaa ggcggcggca ggagaggttg tgggtgctagt ttctctaagc catccagtgc 60  
catcctcgctc gctgcagcga cacacgctct cgccgccgcc atgactgagc agatgaccct 120  
tcgtggcacc ctcaagggcc acaacggctg ggtaaccag atcgctacta ccccgagtt 180  
cccgacatg atcctctccg cctctcgagg tacggactaa gataagacca tcatcatgtg 240  
gaaactgacc aggg atg aga cca act atg gaa ttc cac agc gtg ctc tgc 290  
Met Arg Pro Thr Met Glu Phe His Ser Val Leu Cys  
-25 -20

ggg gtc act ccc act ttg tta gtg atg tgg tta tct cct cag atg gcc 338  
Gly Val Thr Pro Thr Leu Leu Val Met Trp Leu Ser Pro Gln Met Ala  
-15 -10 -5

agt tcg ccc tct cag gct cct ggg atg gaa ccc tgc gcc tct ggg atc 386  
Ser Ser Pro Ser Gln Ala Pro Gly Met Glu Pro Cys Ala Ser Gly Ile  
1 5 10 15

tca caa cgg gca a 399  
Ser Gln Arg Ala  
20

&lt;210&gt; 334

&lt;211&gt; 188

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 33..188

&lt;221&gt; sig\_peptide

&lt;222&gt; 33..131

&lt;223&gt; Von Heijne matrix

score 5.5

seq SLCLLTAVLVLT/FK

&lt;400&gt; 334

aatgaagggt actagaacac ctgcccattcc at atg gga aaa aaa aaa atc tgg 53  
Met Gly Lys Lys Lys Ile Trp  
-30

acc cct agc tca tat ccc atg ccc agt cat aaa cat gta tcc cta tgt 101  
Thr Pro Ser Ser Tyr Pro Met Pro Ser His Lys Val Ser Leu Cys  
-25 -20 -15

ctt cta acg gtt gca gtt tta gtt ctt aca ttt aag tct tta att cat 149

Leu Leu Thr Val Ala Val Leu Val Leu Thr Phe Lys Ser Leu Ile His  
 -10 -5 1 5  
 ttt gag tda att ttt gca tat gag ata ggg gtc cag ggg 188  
 Phe Glu Xaa Ile Phe Ala Tyr Glu Ile Gly Val Gln Gly  
 10 15

<210> 335  
 <211> 115  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 23..115

<221> sig\_peptide  
 <222> 23..94  
 <223> Von Heijne matrix  
 score 5.5  
 seq CPSLLSPISPSQA/CP

<400> 335  
 ccaatacaca tcactcagtg gc atg agc cct gtc ctc tgc ttc cat cgc tgc 52  
 Met Ser Pro Val Leu Cys Phe His Arg Cys  
 -20 -15  
 tcc tgt ccc tcc ctc ctc agc ccc atc tcc cca tcc cag gcc tgt cct 100  
 Ser Cys Pro Ser Leu Leu Ser Pro Ile Ser Pro Ser Gln Ala Cys Pro  
 -10 -5 1  
 gag ccc ctc ctt ggg 115  
 Glu Pro Leu Leu Gly  
 5

<210> 336  
 <211> 300  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 197..298

<221> sig\_peptide  
 <222> 197..268  
 <223> Von Heijne matrix  
 score 5.5  
 seq IMFVCMVCVCVC/VY

<400> 336  
 catgcttggt gtaacgtgtc aaacaatata gaggtgtagg gaaaatacct agtgccaccc 60  
 tccactccaa aaccccatgt cgccagagat aaccatttat tcagacagtg agtatctatt 120  
 aagtatctat tgctaggctt tggagatagc ataatgaaca aaatggatgt gctctctgcc 180  
 cttgtgattt ggacag atg ctt cag tta tct ttt tct gtg ttt ata ttg att 232  
 Met Leu Gln Leu Ser Phe Ser Val Phe Ile Leu Ile  
 -20 -15  
 atg ttt gta tgt atg tgc gtg tgt gtg tgt gtg tat cga ctg 280  
 Met Phe Val Cys Met Cys Val Cys Val Cys Val Cys Val Tyr Arg Leu  
 -10 -5 1  
 ttt tct tcc tcc tcc ccg gg 300  
 Phe Ser Ser Ser Ser Pro  
 5 10

<210> 337  
 <211> 307

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 5..307

&lt;221&gt; sig\_peptide

&lt;222&gt; 5..277

&lt;223&gt; Von Heijne matrix

score 5.5

seq RVLLGAGIPPVSS/AP

&lt;400&gt; 337

```

caca atg aag tcg act gtt tcg tcg agg gaa gtg gcc acc gtt gat aaa      49
    Met Lys Ser Thr Val Ser Ser Arg Glu Val Ala Thr Val Asp Lys
      -90                      -85                      -80
atg aaa aga cgc cat gca gaa tac tgt gca cag ggt ctc cag aga ttt      97
Met Lys Arg Arg His Ala Glu Tyr Cys Ala Gln Gly Leu Gln Arg Phe
      -75                      -70                      -65
aaa gcc caa ctt tct caa gat acc ctt ccc cav cat cca cat ctg gag     145
Lys Ala Gln Leu Ser Gln Asp Thr Leu Pro Xaa His Pro His Leu Glu
      -60                      -55                      -50                      -45
awa gag aag ggg ctt gaa ggc ttg gag gaa aat gtg cct cta aag gga     193
Xaa Glu Lys Gly Leu Glu Gly Leu Glu Glu Asn Val Pro Leu Lys Gly
      -40                      -35                      -30
gag aaa cct gga gaa ggg ggt cca gag tct cct aag aag aga aga agg     241
Glu Lys Pro Gly Glu Gly Gly Pro Glu Ser Pro Lys Lys Arg Arg Arg
      -25                      -20                      -15
gtg ctt ctc gga gcg ggc atc cca cca gta agc tca gct ccc agg aga     289
Val Leu Leu Gly Ala Gly Ile Pro Pro Val Ser Ser Ala Pro Arg Arg
      -10                      -5                      1
cag agc cag cag gca aca
Gln Ser Gln Gln Ala Thr
5                      10

```

&lt;210&gt; 338

&lt;211&gt; 123

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 16..123

&lt;221&gt; sig\_peptide

&lt;222&gt; 16..75

&lt;223&gt; Von Heijne matrix

score 5.5

seq VHLFFFFFFXETGS/RS

&lt;400&gt; 338

```

ttaattaaac tgtgg atg cac aac agt tgt aga cct gtg cac ctt ttt ttc      51
    Met His Asn Ser Cys Arg Pro Val His Leu Phe Phe
      -20                      -15                      -10
ttt ttt ttt yct gag aca ggt tct cgt tct aat ycc tgg ctg gag tsc      99
Phe Phe Phe Xaa Glu Thr Gly Ser Arg Ser Asn Xaa Trp Leu Glu Xaa
      -5                      1                      5
agt ggt gcg atc ata gct aac tcc
Ser Gly Ala Ile Ile Ala Asn Ser
10                      15

```

&lt;210&gt; 339

<211> 451  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 318..449

<221> sig\_peptide  
 <222> 318..443  
 <223> Von Heijne matrix  
 score 5.40000009536743  
 seq TFRLLYLSLPVSA/GP

<221> misc\_feature  
 <222> 310..311,394  
 <223> n=a, g, c or t

<400> 339  
 gtcacaaaag gagcactaag agcctgcttt actttcttcc tcagttgagt cgtggggaca 60  
 gcttgaagga gccaacctca attgcagaga gcagccgtca cccagctac cgctcagagc 120  
 ccagcttgga accagagagc ttccgttctc ctacctttgg caaaagtttt cacttcgatc 180  
 cactatccag tggctcacgc tcctccagcc tcaagtcagc ccagggcaca ggctttgagc 240  
 tgggccagtt gcaatccatt cgttcagagg gcaccacctc cacctcctaa taagagcctg 300  
 gccaacaccagn nacgcaa atg gaa gcc tat ctt aat gac agc ttg ctc aca 350  
 Met Glu Ala Tyr Leu Asn Asp Ser Leu Leu Thr  
 -40 -35  
 cct tca gac agc cct gat ttt gag tca gtg cag gca ggg cct gna gcc 398  
 Pro Ser Asp Ser Pro Asp Phe Glu Ser Val Gln Ala Gly Pro Xaa Ala  
 -30 -25 -20  
 aga ccc acc ttt agg cta tac ctc tcc ctt cct gtc agc cag gct ggc 446  
 Arg Pro Thr Phe Arg Leu Tyr Leu Ser Leu Pro Val Ser Gln Ala Gly  
 -15 -10 -5 1  
 cca gc 451  
 Pro

<210> 340  
 <211> 304  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 94..303

<221> sig\_peptide  
 <222> 94..135  
 <223> Von Heijne matrix  
 score 5.40000009536743  
 seq PALGPALLQGSIX/RV

<221> misc\_feature  
 <222> 244..245  
 <223> n=a, g, c or t

<400> 340  
 gcgcagggga gaaacaaggc gccttggagt tcaggtgact cccacacggg tcatgctgtt 60  
 gtctcctgat ccagccggcc ctgccaggtg acc atg cct gct ctg ggc cca gct 114  
 Met Pro Ala Leu Gly Pro Ala  
 -10  
 ctt ctc cag ggc tct ctg kgc cgv gtg ggt cct cac cct cca gcs cct 162

Leu	Leu	Gln	Gly	Ser	Leu	Xaa	Arg	Val	Gly	Pro	His	Pro	Pro	Ala	Pro	
		-5					1				5					
tcc	acc	aac	tgc	att	cac	tcc	caa	tgg	cac	gta	tct	gca	gca	csk	ggc	210
Ser	Thr	Asn	Cys	Ile	His	Ser	Gln	Trp	His	Val	Ser	Ala	Ala	Xaa	Gly	
10					15					20					25	
aag	gga	ccc	cac	ctc	agg	cac	cct	ctr	sct	ggg	nns	tac	caa	ctt	cct	258
Lys	Gly	Pro	His	Leu	Arg	His	Pro	Leu	Xaa	Gly	Xaa	Tyr	Gln	Leu	Pro	
				30					35					40		
gtt	cca	gct	gag	ccc	tgg	gct	gca	gct	gga	ggc	cac	agt	gtc	cac	c	304
Val	Pro	Ala	Glu	Pro	Trp	Ala	Ala	Ala	Gly	Gly	His	Ser	Val	His		
			45					50					55			

```
<210> 341
<211> 379
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 315..377
```

```
<221> sig_peptide
<222> 315..371
<223> Von Heijne matrix
      score 5.40000009536743
      seq LCCSGCVPSLCCS/SY
```

[illegible]

```
<210> 342
<211> 289
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 223..288
```

```
<221> sig_peptide
<222> 223..264
<223> Von Heijne matrix
      score 5.40000009536743
      seq AHSILLLASQAGC/LR
```

[illegible]

```

-10          -5          1          5          289
aat tgg g
Asn Trp

<210> 343
<211> 169
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 78..167

<221> sig_peptide
<222> 78..137
<223> Von Heijne matrix
      score 5.40000009536743
      seq WVFLVAIFKGVHC/EG

<400> 343
agctctggga gaggagcccc cgccctggga ttcccagggtg ttttcatttg gtgatcagca      60
ctgaacacag aagagtc atg acg gag ttt ggg ctg agc tgg gtt ttc ctt      110
          Met Thr Glu Phe Gly Leu Ser Trp Val Phe Leu
          -20          -15          -10
gtt gct att ttt aaa ggt gtc cac tgt gaa ggt cma att ggt gga gtc      158
Val Ala Ile Phe Lys Gly Val His Cys Glu Gly Xaa Ile Gly Gly Val
          -5          1          5
ggg ggg gcg gg      169
Gly Gly Ala
      10

<210> 344
<211> 112
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 63..110

<221> sig_peptide
<222> 63..104
<223> Von Heijne matrix
      score 5.40000009536743
      seq NTVFLLLFFGCFF/FE

<400> 344
tgtgttttct ctgtcccaaa ttaaattgcat tggggaagtt tataattaca ggaattccac      60
gc atg aac act gtt ttt ttg ttg ttg ttt ttt ggt tgt ttt ttt ttt      107
      Met Asn Thr Val Phe Leu Leu Leu Phe Phe Gly Cys Phe Phe Phe
          -10          -5          1
gag ac      112
Glu

<210> 345
<211> 349
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 207..347

```

&lt;221&gt; sig\_peptide

&lt;222&gt; 207..278

&lt;223&gt; Von Heijne matrix

score 5.40000009536743

seq SCCCLSSSSFIAG/RR

&lt;400&gt; 345

tcacgcgtcta	cgtggacggc	agctggagcc	cgtggagcaa	gtggtcggcc	tgtgggctgg	60
actgcaccca	ctggcggacc	gtgagtgtc	tgaccagca	ccccgcaacg	gaggggagga	120
gtgccagggc	actgacctgg	acaccgcaa	ctgtaccagt	gacctctgtg	tacacactgc	180
ttctggccct	gaggacgtgg	ccctct atg	tgg gcc tca	tcg ccg tgg	ccg tct	233

Met Trp Ala Ser Ser Pro Trp Pro Ser

-20

gcb tgg tcc	tgc tgc tgc	ttg tcc	tca tcc	tcg ttt	att gcc	gga aga	281
Ala Trp Ser	Cys Cys Cys	Leu Ser	Ser Ser	Ser Ser	Phe Ile	Ala Gly Arg	

-15

-10

-5

1

agg agg ggc	tgg act cag	atg tgg	ctg act	cgt cca	ttc tca	cct cag	329
Arg Arg Gly	Trp Thr Gln	Met Trp	Leu Thr	Arg Pro	Phe Ser	Pro Gln	

5

10

15

gct tcc agc	ccg tca	gca tc	349
Ala Ser Ser	Pro Ser	Ala	

20

&lt;210&gt; 346

&lt;211&gt; 191

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 45..191

&lt;221&gt; sig\_peptide

&lt;222&gt; 45..143

&lt;223&gt; Von Heijne matrix

score 5.40000009536743

seq FMLIILSAILLNS/FI

&lt;400&gt; 346

ccttccatag	gtgggtgtacc	gttttgcatt	cccatcagca	ctgt atg	aca atg	ccc	56
				Met Thr	Met Pro		

-30

att tct tca	tat tcc	cag aat	gtg ttg	tca aac	ttt cac	gat ggc	tat	104
Ile Ser Ser	Tyr Ser	Gln Asn	Val Leu	Ser Asn	Phe His	Asp Gly	Tyr	

-25

-20

-15

ttt atg tta	att ata	ctt tct	gcc att	tta cta	aat tct	ttt att	ggg	152
Phe Met Leu	Ile Ile	Leu Ser	Ala Ile	Leu Leu	Asn Ser	Phe Ile	Gly	

-10

-5

1

tgt gtc agc	ttt tat	cat tgc	ttt tct	tgg ggt	tca ggg	191
Cys Val Ser	Phe Tyr	His Cys	Phe Ser	Trp Gly	Ser Gly	

5

10

15

&lt;210&gt; 347

&lt;211&gt; 229

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 144..227

&lt;221&gt; sig\_peptide

&lt;222&gt; 144..203



<223> Von Heijne matrix  
 score 5.40000009536743  
 seq LSLVIFLLTVKHC/FR

<400> 347  
 ttccatataag ccacctcctc ttggtagcca gaagaccctt cggatgatgc cccaggtgta 60  
 aaactctctg gggcccgccc cactcggaag gattactgaa atgagtcatt tccgggacgc 120  
 cttttttact gttgaatgaa agg atg cta aca cat ggg gct tcc ctg tct tta 173  
 Met Leu Thr His Gly Ala Ser Leu Ser Leu  
 -20 -15  
 gtc ata ttt ctg tta aca gtg aag cat tgc ttt aga tac aga gta tac 221  
 Val Ile Phe Leu Leu Thr Val Lys His Cys Phe Arg Tyr Arg Val Tyr  
 -10 -5 1 5  
 aag act tt 229  
 Lys Thr

<210> 348  
 <211> 210  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 106..210

<221> sig\_peptide  
 <222> 106..171  
 <223> Von Heijne matrix  
 score 5.40000009536743  
 seq FWT SIPILPLSSG/RQ

<400> 348  
 aaagaatcca gttgagccta tcgggacttt tgacctacag aactgtgaga taaaaaatgg 60  
 gtgtcgtttt agataaccca tggcagcatt ccctcctctg ctgga atg tgc tca gtg 117  
 Met Ser Ser Val  
 -20  
 gag act gac tgg gga ttc tgg act tcc atc ccc atc ctc cca ctc agc 165  
 Glu Thr Asp Trp Gly Phe Trp Thr Ser Ile Pro Ile Leu Pro Leu Ser  
 -15 -10 -5  
 agt ggt agg cag ctc ccc ctc ccc act aga gaa tgg gga atg tgg 210  
 Ser Gly Arg Gln Leu Pro Leu Pro Thr Arg Glu Trp Gly Met Trp  
 1 5 10

<210> 349  
 <211> 431  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 184..429

<221> sig\_peptide  
 <222> 184..282  
 <223> Von Heijne matrix  
 score 5.40000009536743  
 seq LSAILSM LSLSFS/TT

<221> misc\_feature  
 <222> 214  
 <223> n=a, g, c or t

<400> 349  
aggacatcct ctccaatcca ccacacacca ccttaccctt ctgctggcaa gaggggacct 60  
gattcatcct cacgctaaac actcattcta cccaactgat tgagacagaa cagaagataa 120  
ctgaaacttc tctgccttcc cgctgcaaga agtgaatgag cgatccctct caactgactk 180  
raa atg ttt gcc tca ccc agg aga tgg agc tct ncg aag gcc ttc tct 228  
Met Phe Ala Ser Pro Arg Arg Trp Ser Ser Xaa Lys Ala Phe Ser  
-30 -25 -20  
ggc cag cgg aca ctc cta tct gcc atc ctc agc atg cta tca ctc agc 276  
Gly Gln Arg Thr Leu Leu Ser Ala Ile Leu Ser Met Leu Ser Leu Ser  
-15 -10 -5  
ttc tcc aca aca tcc ctg ctc agc aac tac tgg ttt gtg ggc aca cag 324  
Phe Ser Thr Thr Ser Leu Leu Ser Asn Tyr Trp Phe Val Gly Thr Gln  
1 5 10  
aag gtg ccc aag ccc ctg tgc gag aaa ggt ctg gca gcc aag tgc ttt 372  
Lys Val Pro Lys Pro Leu Cys Glu Lys Gly Leu Ala Ala Lys Cys Phe  
15 20 25 30  
gac atg cca gtg tcc ctg gat gga gat acc aac aca tcc acc cag gag 420  
Asp Met Pro Val Ser Leu Asp Gly Asp Thr Asn Thr Ser Thr Gln Glu  
35 40 45  
gtg gta mma ta 431  
Val Val Xaa

<210> 350  
<211> 386  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 197..385

<221> sig\_peptide  
<222> 197..244  
<223> Von Heijne matrix  
score 5.40000009536743  
seq HSVFLCAPALVFP/RP

<400> 350  
aaagtaaagc ggaggcagcg ggggaagatg gcggcggccg ttccacagcg ggcgtggacc 60  
gtggagcagc tgcgcagtga gcagctgccc aagaaggaca ttatcaagtt tctgcaggaa 120  
cacggttcag attcggtagc agaggcgtag gggcgccggt gctggtgcgg ctgagggacg 180  
cctcaccctt ctggag atg ccc ata cat tcc gta ttc ctc tgt gcc ccc gcc 232  
Met Pro Ile His Ser Val Phe Leu Cys Ala Pro Ala  
-15 -10 -5  
ctc gtc ttc ccg cgg ccg gtg gcc tgg aag gcg gag agg ccc agc ttg 280  
Leu Val Phe Pro Arg Pro Val Ala Trp Lys Ala Glu Arg Pro Ser Leu  
1 5 10  
tgc ttt ggt gcc tgc ctc ccg cct ctc ggg cgt tct cta ctg ggg cag 328  
Cys Phe Gly Ala Ser Leu Pro Pro Leu Gly Arg Ser Leu Leu Gly Gln  
15 20 25  
ggg agc agc ttt att tct tgg ggc aca cag gct gca att gta gag tta 376  
Gly Ser Ser Phe Ile Ser Trp Gly Thr Gln Ala Ala Ile Val Glu Leu  
30 35 40  
kaa cct cat t 386  
Xaa Pro His  
45

<210> 351  
<211> 307  
<212> DNA  
<213> Homo sapiens

<220>

<221> CDS  
<222> 68..307

<221> sig\_peptide  
<222> 68..253  
<223> Von Heijne matrix  
score 5.30000019073486  
seq LACVFFLSHPLFX/XP

<221> misc\_feature  
<222> 279  
<223> n=a, g, c or t

<400> 351  
ttttactctg taattgttac taattgattt ttgmataggg agcacattcc catggttcaa 60  
aattcaa atg gta tac gat gaa aaa tct ctc tcc tgt tcc cat acc cca 109  
Met Val Tyr Asp Glu Lys Ser Leu Ser Cys Ser His Thr Pro  
-60 -55 -50  
gcc acc cag ttc ctc tcc tgg gat gca tcc agt gtt tac agt ttc tta 157  
Ala Thr Gln Phe Leu Ser Trp Asp Ala Ser Ser Val Tyr Ser Phe Leu  
-45 -40 -35  
tat atc ctc tca gca aga gtt aat gta gac gta dgc agm tac att cgt 205  
Tyr Ile Leu Ser Ala Arg Val Asn Val Asp Val Xaa Xaa Tyr Ile Arg  
-30 -25 -20  
gtg tac ata ctt gcc tgt gtg ttt ttc ctc tca cac ccc ctt ttt aad 253  
Val Tyr Ile Leu Ala Cys Val Phe Phe Leu Ser His Pro Leu Phe Xaa  
-15 -10 -5  
sra cca aat ggt agt gta tat tgt cnm cgt cat tct ccc cct tac ctt 301  
Xaa Pro Asn Gly Ser Val Tyr Cys Xaa Arg His Ser Pro Pro Tyr Leu  
1 5 10 15  
ttt tgc 307  
Phe Cys

<210> 352  
<211> 170  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 56..169  
  
<221> sig\_peptide  
<222> 56..163  
<223> Von Heijne matrix  
score 5.30000019073486  
seq VCLLIISLVLISG/LG

<400> 352  
gttcttttggg gatacaaaaca ctgtattttg agtaatcttt tccctatatatt tcgaa atg 58  
Met  
ctg cct tta tca cct act aaa ttc cta aat gtg ttc ttg ggc ctg ttc 106  
Leu Pro Leu Ser Pro Thr Lys Phe Leu Asn Val Phe Leu Gly Leu Phe  
-35 -30 -25 -20  
ctc tat tat ctt caa ttg gta tgt ctg ctt att att tct ttg gtt ttg 154  
Leu Tyr Tyr Leu Gln Leu Val Cys Leu Leu Ile Ile Ser Leu Val Leu  
-15 -10 -5  
ata tct ggg tta ggg g 170  
Ile Ser Gly Leu Gly  
1

<210> 353

194

<211> 293  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 149..292

<221> sig\_peptide  
 <222> 149..235  
 <223> Von Heijne matrix  
 score 5.30000019073486  
 seq LNQTLMLLREVLVLA/SH

<400> 353  
 tttctaattct sbtcaaattt tatcaccata caatcagtggt taktgttgga aatagtgcaa 60  
 ctgcattatt gactaccatt gaagaaatgc atttgctaag caaaaaaata ttcttcaatt 120  
 agcttgaagt cttcatgcaa gttaaatta atg gac aag gtt gaa ctc cca cca 172  
 Met Asp Lys Val Glu Leu Pro Pro  
 -25  
 cct gat ctt gga cca agt tct gca cta aat cag aca ctc atg ttg ctg 220  
 Pro Asp Leu Gly Pro Ser Ser Ala Leu Asn Gln Thr Leu Met Leu Leu  
 -20 -15 -10  
 cgt gaa gtt tta gca tct cac gat tct tca gtk gta cca tta gat gct 268  
 Arg Glu Val Leu Ala Ser His Asp Ser Ser Val Val Pro Leu Asp Ala  
 -5 1 5 10  
 cgt caa gct gat ttt gtg cag ggg g 293  
 Arg Gln Ala Asp Phe Val Gln Gly  
 15

<210> 354  
 <211> 331  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 148..330

<221> sig\_peptide  
 <222> 148..243  
 <223> Von Heijne matrix  
 score 5.30000019073486  
 seq LVWLWFVVPQTIT/MI

<221> misc\_feature  
 <222> 124  
 <223> n=a, g, c or t

<400> 354  
 catttctagc ttttgdktta aagtgcacaga cttgccactc ttcttttccc ttgaacactt 60  
 acaggctgtg ggagggttat tagttggtct aatttcaata mtgttccttt cyccagggaa 120  
 ttgnraggcc caaggagagg gagagag atg ggg gga aca gct ggt tgg agc agt 174  
 Met Gly Gly Thr Ala Gly Trp Ser Ser  
 -30 -25  
 cag aac aca cac aac att kga gta cac cat ctt gtg tgg ctg tgg ttc 222  
 Gln Asn Thr His Asn Ile Xaa Val His His Leu Val Trp Leu Trp Phe  
 -20 -15 -10  
 gtg gtc ccc caa aca att aca atg ata aca cca aag atc act gaa cac 270  
 Val Val Pro Gln Thr Ile Thr Met Ile Thr Pro Lys Ile Thr Glu His  
 -5 1 5  
 aga cca sta ata aca gat atr dtr ata atg aya aca ttt gaa awa ttg 318

195

Arg Pro Xaa Ile Thr Asp Xaa Xaa Ile Met Xaa Thr Phe Glu Xaa Leu  
 10 15 20 25  
 gga gaa tta ccc a  
 Gly Glu Leu Pro

331

<210> 355  
 <211> 93  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 2..91

<221> sig\_peptide  
 <222> 2..55  
 <223> Von Heijne matrix  
 score 5.30000019073486  
 seq ALYLCVCVCVCLI/AR

<400> 355  
 t atg tgt ctv agt gta gct ttg tat tta tgt gtg tgt gtg tgt gta tgt 49  
 Met Cys Leu Ser Val Ala Leu Tyr Leu Cys Val Cys Val Cys Val Cys  
 -15 -10 -5  
 ctg att gca cgg gtg tac ttt tgt att tat gtg tgt gtg tgg tt 93  
 Leu Ile Ala Arg Val Tyr Phe Cys Ile Tyr Val Cys Val Trp  
 1 5 10

<210> 356  
 <211> 178  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 92..178

<221> sig\_peptide  
 <222> 92..133  
 <223> Von Heijne matrix  
 score 5.30000019073486  
 seq LHLLFGLFPVLWM/FL

<400> 356  
 tgacccttgt ccagtctttt ccaggaaaaa catgccctca agatgttttt ctatcttgag 60  
 gaaatgatgg aaatgagata gttccaaggg t atg ctt cac ctt ctt ttt ggc 112  
 Met Leu His Leu Leu Phe Gly  
 -10  
 tta ttt cct gtt ctt tgg atg ttt cta gtg tat ttc ttt ctt tct tct 160  
 Leu Phe Pro Val Leu Trp Met Phe Leu Val Tyr Phe Phe Leu Ser Ser  
 -5 1 5  
 ttt ttt ttt ttt ttt ttt 178  
 Phe Phe Phe Phe Phe Phe  
 10 15

<210> 357  
 <211> 107  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 40..105

<221> sig\_peptide  
 <222> 40..93  
 <223> Von Heijne matrix  
 score 5.30000019073486  
 seq CVYLFCACMCVCA/FF

<221> misc\_feature  
 <222> 54  
 <223> n=a, g, c or t

<400> 357  
 tatatttata taaatatata taaatacaca catatatat atg tat gtg tgt atn 54  
 Met Tyr Val Cys Xaa  
 -15  
 tgt gtg tat ctt ttt tgt gca tgt atg tgt gta tgt gct ttt ttt ttt 102  
 Cys Val Tyr Leu Phe Cys Ala Cys Met Cys Val Cys Ala Phe Phe Phe  
 -10 -5 1  
 ttt tt 107  
 Phe

<210> 358  
 <211> 209  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 44..208

<221> sig\_peptide  
 <222> 44..151  
 <223> Von Heijne matrix  
 score 5.30000019073486  
 seq FLFTLIGASLLQS/AS

<400> 358  
 ggggggagac gtaagggcgc tccgcgagcc cgtctctcct cga atg aaa sga aac 55  
 Met Lys Xaa Asn  
 -35  
 aac ctc cgg cga cag agc ccc gct ctc agg cac tgc tgg aga mcc gag 103  
 Asn Leu Arg Arg Gln Ser Pro Ala Leu Arg His Cys Trp Arg Xaa Glu  
 -30 -25 -20  
 acc gac ttc ttt ctc ttt acc ctc att ggc gct tct ctc ctg cag tcc 151  
 Thr Asp Phe Phe Leu Phe Thr Leu Ile Gly Ala Ser Leu Leu Gln Ser  
 -15 -10 -5  
 gcc tct ggg ccc tgc cgc att tct tsa smc tta aag tgg cat tct aaa 199  
 Ala Ser Gly Pro Cys Arg Ile Ser Xaa Xaa Leu Lys Trp His Ser Lys  
 1 5 10 15  
 ggc act tta a 209  
 Gly Thr Leu

<210> 359  
 <211> 298  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 135..296

<221> sig\_peptide

&lt;222&gt; 135..194

&lt;223&gt; Von Heijne matrix

score 5.30000019073486

seq LGLGLPLLPPNHP/SV

&lt;400&gt; 359

```

agcatttgcm ttcggagggc cagargggca ggcagagctt aattccttgg gcaaggctgg      60
ggctgttgga atgggggtctg gaggccagga gccaccctgt ctgggccaga aaggggcctk      120
ggtgcagggc aggc atg tgg ccc aag arg ggg cta ctg gga ttg ggg ctc      170
          Met Trp Pro Lys Xaa Gly Leu Leu Gly Leu Gly Leu
          -20          -15          -10
cca ctg ctg ccc cct aac cat ccc tcg gta gcc caa ggg aca ctc gtt      218
Pro Leu Leu Pro Pro Asn His Pro Ser Val Ala Gln Gly Thr Leu Val
          -5          1          5
tcc tcc cac tct ggt tct ggc tct gag ggt agg gtg gcg ctc agg agt      266
Ser Ser His Ser Gly Ser Gly Ser Glu Gly Arg Val Ala Leu Arg Ser
          10          15          20
gat gtc cac agc ccc aag aca acc csc caa cg      298
Asp Val His Ser Pro Lys Thr Thr Xaa Gln
25          30

```

&lt;210&gt; 360

&lt;211&gt; 460

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 54..458

&lt;221&gt; sig\_peptide

&lt;222&gt; 54..179

&lt;223&gt; Von Heijne matrix

score 5.30000019073486

seq AMAXLFLSAPPQA/EV

&lt;221&gt; misc\_feature

&lt;222&gt; 150,285,328

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 360

```

gaggttgggc tgccgtgctg ctcggcgggcg ctgaggccaa atagttgcat cac atg      56
                                     Met
tat cta atc cga gag tct cat gct tct ggt agc tcc tca gtg acc agc      104
Tyr Leu Ile Arg Glu Ser His Ala Ser Gly Ser Ser Ser Val Thr Ser
          -40          -35          -30
tcc tgc tca ctg mcc tca gra agc ccc aac cct cag gca atg gck ncc      152
Ser Cys Ser Leu Xaa Ser Xaa Ser Pro Asn Pro Gln Ala Met Ala Xaa
          -25          -20          -15          -10
ttg ttc ctg tct gcc cca ccc cag gcc gag gtg acc ttc gag gac gtg      200
Leu Phe Leu Ser Ala Pro Pro Gln Ala Glu Val Thr Phe Glu Asp Val
          -5          1          5
gct gtg tac ctc tcc cgg gag gaa tgg ggc cgc ctg ggc cct gct cag      248
Ala Val Tyr Leu Ser Arg Glu Glu Trp Gly Arg Leu Gly Pro Ala Gln
          10          15          20
agg ggc bkc tac agg gac gtg atg ctg gag acc tac ngg aac bta gtc      296
Arg Gly Xaa Tyr Arg Asp Val Met Leu Glu Thr Tyr Xaa Asn Xaa Val
          25          30          35
tca ctg gga gta gga cct gca ggc ccc aag cnt gga gtg atc tcg cag      344
Ser Leu Gly Val Gly Pro Ala Gly Pro Lys Xaa Gly Val Ile Ser Gln
          40          45          50          55
ttg gag cga ggg gat gag ccc tgg gtc ctg gat gtt cag ggc acc tct      392

```

198

Leu Glu Arg Gly Asp Glu Pro Trp Val Leu Asp Val Gln Gly Thr Ser  
 60 65 70  
 ggg aaa gag cac ctg aag aag tca aca gcc cag ctc ttg gga cca gaa 440  
 Gly Lys Glu His Leu Lys Lys Ser Thr Ala Gln Leu Leu Gly Pro Glu  
 75 80 85  
 ctg aag tac aag gag ttg ay 460  
 Leu Lys Tyr Lys Glu Leu  
 90

<210> 361  
 <211> 318  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 153..317

<221> sig\_peptide  
 <222> 153..263  
 <223> Von Heijne matrix  
 score 5.30000019073486  
 seq ALSSLCVSWGTTSS/TV

<400> 361  
 ctctttttccg gttaacgcgg cgtgagaagc catgagcagc aaagtctctc gcgacaccct 60  
 gtacgaggcg gtgcgggaag tcctgcacgg gaaccagcgc aasgccgcaa gttcctggag 120  
 acggtggagt tgcaggatca gcttgaagaa ct atg atc ccc aga agg aca agc 173  
 Met Ile Pro Arg Arg Thr Ser  
 -35  
 gct tct cgg gca ccg tca gtc ccc caa aac gca ggc tta agt cca ctc 221  
 Ala Ser Arg Ala Pro Ser Val Pro Gln Asn Ala Gly Leu Ser Pro Leu  
 -30 -25 -20 -15  
 ccc gcc cta agt tct ctg tgt gtg tcc tgg ggg acc agc agc act gtg 269  
 Pro Ala Leu Ser Ser Leu Cys Val Ser Trp Gly Thr Ser Ser Thr Val  
 -10 -5 1  
 acg agg cta agg ccg tgg ata tcc ccc aca tgg aca tcg agg gcg cgg g 318  
 Thr Arg Leu Arg Pro Trp Ile Ser Pro Thr Trp Thr Ser Arg Ala Arg  
 5 10 15

<210> 362  
 <211> 360  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 192..359

<221> sig\_peptide  
 <222> 192..233  
 <223> Von Heijne matrix  
 score 5.30000019073486  
 seq VCIFCFLTSKAFP/NP

<221> misc\_feature  
 <222> 277  
 <223> n=a, g, c or t

<400> 362  
 tattgggttg ttttcttatt atcaaattgt gaaagttctt tacatattct gggtagaact 60  
 cctttatcag atacatgttt tgcaaattgt ttctaccatt ctctgtctdh tctttctctt 120



199

```

aatactttca cagtttttca tagcagaaat ttataaatta atgaagccca ctttatactt 180
ttattttttt t atg gtt tgc atc ttt tgt ttc tta act tcg aaa gct ttt 230
      Met Val Cys Ile Phe Cys Phe Leu Thr Ser Lys Ala Phe
                -10                -5
cct aac cct aga tca cag gat ttt ctc tta gat ttc tct agg cat tnt 278
Pro Asn Pro Arg Ser Gln Asp Phe Leu Leu Asp Phe Ser Arg His Xaa
      1                5                10                15
ata ggt tta ggt ttc aca ttt agg tcc gca atg cat ttt gaa aac ttc 326
Ile Gly Leu Gly Phe Thr Phe Arg Ser Ala Met His Phe Glu Asn Phe
                20                25                30
cgt ctg waa ggt ttg ggt caa gat tcc ctt tgt c 360
Arg Leu Xaa Gly Leu Gly Gln Asp Ser Leu Cys
      35                40

```

<210> 363  
 <211> 212  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 138..212

<221> sig\_peptide  
 <222> 138..197  
 <223> Von Heijne matrix  
       score 5.30000019073486  
       seq GFCSVTSSPLASA/GR

<221> misc\_feature  
 <222> 152  
 <223> n=a, g, c or t

```

<400> 363
cacaaaaatca aaaackkagt tgacgtatgc cactttccag ttactattga gatatatatg 60
cgtgtgtgta tatattacat atatatgtta tatatcatat tkatatattt akaaawttat 120
atmgavcata catatat atg taw atr tat ktn kkt ava ggg ttt tgc tct 170
      Met Xaa Xaa Tyr Xaa Xaa Xaa Gly Phe Cys Ser
                -20                -15                -10
gtc aca agc agt cct ctt gcc tca gca ggt agg act aca cgc 212
Val Thr Ser Ser Pro Leu Ala Ser Ala Gly Arg Thr Thr Arg
      -5                1                5

```

<210> 364  
 <211> 242  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 127..240

<221> sig\_peptide  
 <222> 127..195  
 <223> Von Heijne matrix  
       score 5.30000019073486  
       seq LVPCPLLISSVALS/VK

<221> misc\_feature  
 <222> 71,73  
 <223> n=a, g, c or t

<400> 364  
 actttaactt cctggagctc taatttctcc ttctcaggtc gaagaatgcc attcactccc 60  
 aagtggtag nmncagcag ccagtggttag gaagggtcat caagtcagtt gtcagaaacc 120  
 tcactk atg tca ctg twt ahg cta tgt gac cct gac cta gtt cct tgc 168  
     Met Ser Leu Xaa Xaa Leu Cys Asp Pro Asp Leu Val Pro Cys  
         -20                      -15                      -10  
 cct ctc ttg atc tca gtt gct tta tct gta aaa ttt cac att tkt cag 216  
 Pro Leu Leu Ile Ser Val Ala Leu Ser Val Lys Phe His Ile Xaa Gln  
         -5                      1                      5  
 caa gtc aac ctt cca tgt tcc tct ca 242  
 Gln Val Asn Leu Pro Cys Ser Ser  
         10                      15

<210> 365  
 <211> 248  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 7..246

<221> sig\_peptide  
 <222> 7..123  
 <223> Von Heijne matrix  
     score 5.30000019073486  
     seq LXPCLTSFSCXGA/SF

<400> 365  
 tgtaca atg atg atc ctt atc cta att ctt gag cat atc gtc acc kcc 48  
     Met Met Ile Leu Ile Leu Ile Leu Glu His Ile Val Thr Xaa  
         -35                      -30  
 aaa aga aac ccc aaa cct gtt aca gtc cct gct ttt ctg csc cct tgc 96  
 Lys Arg Asn Pro Lys Pro Val Thr Val Pro Ala Phe Leu Xaa Pro Cys  
 -25                      -20                      -15                      -10  
 ttg act tct ttc tct tgt kct gga gca tct ttc tct ctk ttw ggt gdg 144  
 Leu Thr Ser Phe Ser Cys Xaa Gly Ala Ser Phe Ser Leu Xaa Gly Xaa  
         -5                      1                      5  
 aga agg ggt tgg caa cat ggc agc tgc tgc tcc acc att ccc tta ttt 192  
 Arg Arg Gly Trp Gln His Gly Ser Cys Cys Ser Thr Ile Pro Leu Phe  
         10                      15                      20  
 csa act cta aat tcc ctt ggg cag gga ctc att ggc cca gcc tac ata 240  
 Xaa Thr Leu Asn Ser Leu Gly Gln Gly Leu Ile Gly Pro Ala Tyr Ile  
         25                      30                      35  
 ggt gcd gg 248  
 Gly Ala  
 40

<210> 366  
 <211> 351  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 293..349

<221> sig\_peptide  
 <222> 293..340  
 <223> Von Heijne matrix  
     score 5.30000019073486  
     seq HAISILLICIGASS/QG

<221> misc\_feature  
 <222> 36  
 <223> n=a, g, c or t

<400> 366  
 aaaacatata tccacacaaa aacttgcaca cataknttca tagcagcatt attcatccaa 60  
 aaagtagagg tactcaaag actttcaact gataaacaca gatgaacaaa atgtatgtcc 120  
 aaacagtaga atattattca gctataaaaa agaacagagt acacttagca aactaagaat 180  
 agaaggaaact tectcaatct gataaaggac atccatgaaa aaccaccac taatgtcata 240  
 cttaatcatg aaaaaccgaa tgcttttctc ctaagatagg aaaaagacaa gt atg tct 298  
 Met Ser  
 -15  
 act cat gcc atc tct att cta ctt tgt att ggt gct tct agc cag ggc 346  
 Thr His Ala Ile Ser Ile Leu Leu Cys Ile Gly Ala Ser Ser Gln Gly  
 -10 -5 1  
 agg gg 351  
 Arg

<210> 367  
 <211> 208  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 7..207  
 <221> sig\_peptide  
 <222> 7..99  
 <223> Von Heijne matrix  
 score 5.19999980926514  
 seq ATVNAAASLPPCFG/VK

<400> 367  
 gtctcg atg gag gag caa gaa acg gaa gag gtc ggg ggg aga agc agc 48  
 Met Glu Glu Gln Glu Thr Glu Glu Val Gly Gly Arg Ser Ser  
 -30 -25 -20  
 cgg aaa aat gca gcc acc gtc aac gcc gcc tcc ctg cca ccg tgc ttc 96  
 Arg Lys Asn Ala Ala Thr Val Asn Ala Ala Ser Leu Pro Pro Cys Phe  
 -15 -10 -5  
 ggg gta aaa agc tgc cgt tgc cgt cgg tgc agt tgc cgt cgc tgc ctc 144  
 Gly Val Lys Ser Cys Arg Cys Arg Arg Cys Ser Cys Arg Arg Cys Leu  
 1 5 10 15  
 cta tac ttc tct tgg cct cgg gga agg att tcc cca ccg gtg gga caa 192  
 Leu Tyr Phe Ser Trp Pro Arg Gly Arg Ile Ser Pro Pro Val Gly Gln  
 20 25 30  
 tgt gcg ggg agg gga t 208  
 Cys Ala Gly Arg Gly  
 35

<210> 368  
 <211> 446  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 11..445

<221> sig\_peptide  
 <222> 11..109

&lt;223&gt; Von Heijne matrix

score 5.19999980926514

seq CCCHAGASSGATA/WE

&lt;400&gt; 368

```

agaatccaag atg cgc ggg atc car gca aar ggg tct ccg ggc cag agt      49
      Met Arg Gly Ile Gln Ala Lys Gly Ser Pro Gly Gln Ser
                -30                      -25
tcg gcc gst gtt ctg wcg cct tgc tgc tgt cac gcg ggc gct tcg tcc      97
Ser Ala Xaa Val Leu Xaa Pro Cys Cys Cys His Ala Gly Ala Ser Ser
-20                      -15                      -10                      -5
ggg gcg acg gcg tgg gag gag acc ccg cgg tcg cgt tgc cac atc gcc      145
Gly Ala Thr Ala Trp Glu Glu Thr Pro Arg Ser Arg Cys His Ile Ala
                1                5                10
gtt kcg agt aca aat aca gct tca agg ggc cgc acc tgg tgc aga gcg      193
Val Xaa Ser Thr Asn Thr Ala Ser Arg Gly Arg Thr Trp Cys Arg Ala
                15                20                25
acg gga ccg tgc cct tct ggg ccc acg cgg gga gta agc cgg agc aga      241
Thr Gly Pro Cys Pro Ser Gly Pro Thr Arg Gly Val Ser Arg Ser Arg
                30                35                40
ggg ctg ggg gcc ggg ttc ctc tcc ccc ttc tgc tgc ctc ttc gcc ttt      289
Gly Leu Gly Ala Gly Phe Leu Ser Pro Phe Cys Cys Leu Phe Ala Phe
45                50                55                60
cat ccg cgg cta ccc tgg tgt gct gag gtt ccc gtt cca gca gct gca      337
His Pro Arg Leu Pro Trp Cys Ala Glu Val Pro Val Pro Ala Ala Ala
                65                70                75
cac cat atg cgc tgt gga ggg gac ctc ctg gca gcc cct ccg ccg ggt      385
His His Met Arg Cys Gly Gly Asp Leu Leu Ala Ala Pro Pro Pro Gly
                80                85                90
ccc tcc tgg ttc gca cgg ttc cct ccg ctt gtc ccc gag tct ttc cct      433
Pro Ser Trp Phe Ala Arg Phe Pro Pro Leu Val Pro Glu Ser Phe Pro
                95                100                105
cac cat tct gtt c
His His Ser Val
                110

```

&lt;210&gt; 369

&lt;211&gt; 125

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 22..123

&lt;221&gt; sig\_peptide

&lt;222&gt; 22..93

&lt;223&gt; Von Heijne matrix

score 5.19999980926514

seq LIWVFGLSVLSX/FL

&lt;400&gt; 369

```

ctatcaagag gctttccccc t atg ttt tct tct agg agt ttt atg gtt tca      51
      Met Phe Ser Ser Arg Ser Phe Met Val Ser
                -20                      -15
ggg ctt att tgg gtc ttt ggt ctt gta tct gtt ttg agt bga ttt ttg      99
Gly Leu Ile Trp Val Phe Gly Leu Val Ser Val Leu Ser Xaa Phe Leu
                -10                      -5                1
tgt atg gtg tat gat cag ggt cag gg      125
Cys Met Val Tyr Asp Gln Gly Gln
                5                10

```

&lt;210&gt; 370

<211> 132  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 39..131

<221> sig\_peptide  
 <222> 39..77  
 <223> Von Heijne matrix  
 score 5.19999980926514  
 seq MLLAVSLSLVSNCFNF

<400> 370  
 atcttagagg aaagtctttc agtttttccc cattcagt atg tta tta gct gtg agc 56  
 Met Leu Leu Ala Val Ser  
 -10  
 ctg tcc ctt gtc tct aat tgt aac ttt gta ctc act gac caa ctt ttc 104  
 Leu Ser Leu Val Ser Asn Cys Asn Phe Val Leu Thr Asp Gln Leu Phe  
 -5 1 5  
 cct gcc cct gcc tcc ctc atc ccc gaa g 132  
 Pro Ala Pro Ala Ser Leu Ile Pro Glu  
 10 15

<210> 371  
 <211> 127  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 4..126

<221> sig\_peptide  
 <222> 4..90  
 <223> Von Heijne matrix  
 score 5.19999980926514  
 seq TGVFLFSIIGSFG/FP

<400> 371  
 tga atg aac caa gat ttc aac cca gaa att gag gct tca cca caa gtg 48  
 Met Asn Gln Asp Phe Asn Pro Glu Ile Glu Ala Ser Pro Gln Val  
 -25 -20 -15  
 aag act ggg gtt ttc ttg ttt tca att att ggg agt ttt gga ttt cca 96  
 Lys Thr Gly Val Phe Leu Phe Ser Ile Ile Gly Ser Phe Gly Phe Pro  
 -10 -5 1  
 gga atg tgc aat tgt aaa aac cca gcc cgg g 127  
 Gly Met Cys Asn Cys Lys Asn Pro Ala Arg  
 5 10

<210> 372  
 <211> 196  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 125..196

<221> sig\_peptide  
 <222> 125..184  
 <223> Von Heijne matrix

score 5.19999980926514  
seq IVSSLFSWLLSLT/SV

<221> misc\_feature  
<222> 119  
<223> n=a, g, c or t

<400> 372  
taaaaatctt ttatgttcta cccactcctt cctcggtccc tctccccact cctccctccc 60  
cccatcttaa gcccatggca acccctgata tttttactgt ctccatcggt ttgccttbnc 120  
caga atg cca tgt agt tgg agt cat ata gta agt agc ctt ttc agt tgg 169  
Met Pro Cys Ser Trp Ser His Ile Val Ser Ser Leu Phe Ser Trp  
-20 -15 -10  
ctt ctt tca ctt acc agt gtg ccc ggg 196  
Leu Leu Ser Leu Thr Ser Val Pro Gly  
-5 1

<210> 373  
<211> 148  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 56..148

<221> sig\_peptide  
<222> 56..139  
<223> Von Heijne matrix  
score 5.19999980926514  
seq PVLSCCCLTAGRA/RL

<400> 373  
actttcttca caccaggac gcagggtgcc gctgccggcc acagaaaccc caaga atg 58  
Met  
ttt ttc ttt ggc tat tca gag gac atc tat tgt gtg tca ggc cct gtg 106  
Phe Phe Phe Gly Tyr Ser Glu Asp Ile Tyr Cys Val Ser Gly Pro Val  
-25 -20 -15  
ctg agc tgt tgt tgc ctg aca gca gga aga gcg cgg ctc tgg 148  
Leu Ser Cys Cys Cys Leu Thr Ala Gly Arg Ala Arg Leu Trp  
-10 -5 1

<210> 374  
<211> 200  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> 26..199

<221> sig\_peptide  
<222> 26..73  
<223> Von Heijne matrix  
score 5.19999980926514  
seq AALICPWSSQVPS/SP

<400> 374  
ctagggagga ctcaatgctc tttgt atg cct tat gca gcg ctg atc tgt ccc 52  
Met Pro Tyr Ala Ala Leu Ile Cys Pro  
-15 -10  
tgg agt tcc cag gtt ccc agc tcc ccc cct gca agc ctt gaa gcc tcc 100

205

Trp Ser Ser Gln Val Pro Ser Ser Pro Pro Ala Ser Leu Glu Ala Ser  
 -5 1 5  
 agc aac gtc tat ctc cag gag agc agg gca gcc tat gca agt gtt ccg 148  
 Ser Asn Val Tyr Leu Gln Glu Ser Arg Ala Ala Tyr Ala Ser Val Pro  
 10 15 20 25  
 gca gga cca gaa gtg gcc act caa cac acg tcc tca cca gtc acc cct 196  
 Ala Gly Pro Glu Val Ala Thr Gln His Thr Ser Ser Pro Val Thr Pro  
 30 35 40  
 atg g 200  
 Met

<210> 375  
 <211> 112  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 52..111  
 <221> sig\_peptide  
 <222> 52..105  
 <223> Von Heijne matrix  
 score 5.19999980926514  
 seq LTYSLAFLFIKA/GT

<400> 375  
 aataaccctt tcacagcact tgcctgtttt taatgaatct aattattcac a atg caa 57  
 Met Gln  
 ctt tta tat tta aca tac tct tta gct ttc ctg cta ttt atc aag gct 105  
 Leu Leu Tyr Leu Thr Tyr Ser Leu Ala Phe Leu Leu Phe Ile Lys Ala  
 -15 -10 -5  
 ggc acc g 112  
 Gly Thr  
 1

<210> 376  
 <211> 146  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 74..145  
 <221> sig\_peptide  
 <222> 74..133  
 <223> Von Heijne matrix  
 score 5.19999980926514  
 seq AAAVTSSAAPSRA/RQ

<400> 376  
 ggctggagcg cgcgcctcct agcggascgg ggcaattgga aggccgcgcc tcaggaaaac 60  
 aggatggtag tga atg gca ccg agc cgc ccc agg gct gcc gcc gtc acc 109  
 Met Ala Pro Ser Arg Pro Arg Ala Ala Val Thr  
 -20 -15 -10  
 tcc tcg gcg gct ccg agt cgt gcg agg cag ggg gcc c 146  
 Ser Ser Ala Ala Pro Ser Arg Ala Arg Gln Gly Ala  
 -5 1

<210> 377  
 <211> 389  
 <212> DNA

<213> Homo sapiens

**<220>**

&lt;221&gt; CDS

<222> 218..388

```
<221> sig_peptide
```

<222> 218..343

## <223> Von Heijne matrix

```
score 5.19999980926514
```

seq QHLLSWAXQXGRX/QV

```
<221> misc_feature
```

&lt;222&gt; 139

<223> n=a, g, c or t

<400> 377

cattttgtcg	gtagaggcag	aaggwgaagg	tcgggttgta	gaagctgggg	tggccggcag	60
ctcgtctcatc	ggtgttcgtg	ggctttgtcg	gtccgtgcct	cgtctctccc	tggaaaggga	120
gggaggcttc	gacgtcgrnr	aggragmmgc	tgccgcgtta	gttccgagct	tgaagtcact	180
aggactttctc	tcaaacttgt	gtgctgagga	gactcag	atg ttg gcc tca gct cct		235
				Met Leu Ala Ser Ala Pro		
				-40		
agg ctg aac tca gca gat cgg ccc atg aaa act tct gta ttg aga caa						283
Arg Leu Asn Ser Ala Asp Arg Pro Met Lys Thr Ser Val Leu Arg Gln						
-35		-30		-25		
agg aag gga tct gtc aga aag caa cac ttg tta tct tgg gct tdg cag						331
Arg Lys Gly Ser Val Arg Lys Gln His Leu Ser Trp Ala Xaa Gln						
-20		-15		-10		-5
yaa ggh aga kga cag gta gtg gag atc ctg caa tct gaa aag cag act						379
Xaa Gly Arg Xaa Gln Val Val Glu Ile Leu Gln Ser Glu Lys Gln Thr						
	1		5		10	
daa rgt gac g						389
Xaa Xaa Asp						
15						

<210> 378

<211> 143

<212> DNA

<213> Homo sapiens

**<220>**

<221> CDS

<222> 2..142

```
<221> sig_peptide
```

<222> 2..115

## <223> Von Heijne matrix

score 5.19999980926514

seq LHGSLDAVSQAQG/RP

<400> 378

a	atg	tac	ccc	cta	ggc	agg	gga	gag	cag	ggc	cct	gct	gca	ccc	aag	tcc	49
	Met	Tyr	Pro	Leu	Gly	Arg	Gly	Glu	Gln	Gly	Pro	Ala	Ala	Pro	Lys	Ser	
				-35					-30					-25			
tgg	ttg	ctc	ctc	ccc	acc	aca	ctg	gcc	ctc	cat	gga	agc	ctt	gat	gca	97	
Trp	Leu	Leu	Leu	Pro	Thr	Thr	Leu	Ala	Leu	His	Gly	Ser	Leu	Asp	Ala		
		-20					-15				-10						
gtg	agc	cag	gcc	caa	gga	cgc	ccc	ggc	cac	cct	gac	gca	ccc	ccc	a	143	
Val	Ser	Gln	Ala	Gln	Gly	Arg	Pro	Gly	His	Pro	Asp	Ala	Pro	Pro			
	-5					1			5								



<210> 379  
 <211> 261  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 198..260

<221> sig\_peptide  
 <222> 198..245  
 <223> Von Heijne matrix  
 score 5.19999980926514  
 seq FIAALFTIAETWN/QP

<400> 379  
 cagatggtgg tgagggtgta gagaaaaagg aacgcttata cactggtggt gcgagtgttaa 60  
 attagtttaa ccattgtgga agatgatatg gcaattccac aaagacctaa agtcagraat 120  
 tmcattcaaa cccagtaatc ccattactgg gtatatactc aaaggaatat aaattgttgt 180  
 gttacaaaga cacatgc atg cgt gtg ttc att gca gca ctg ttc aca ata 230  
 Met Arg Val Phe Ile Ala Ala Leu Phe Thr Ile  
 -15 -10  
 gca gag aca tgg aat caa ccc aaa tgc cca g 261  
 Ala Glu Thr Trp Asn Gln Pro Lys Cys Pro  
 -5 1 5

<210> 380  
 <211> 228  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 63..227

<221> sig\_peptide  
 <222> 63..152  
 <223> Von Heijne matrix  
 score 5.19999980926514  
 seq LCFLSVHFRRLRWG/DS

<400> 380  
 gggacgtggg aaaatgacta cgcgtcactc gtgatgtcgc gcatccgata ggcccttttc 60  
 ag atg gca aaa ggc ctg agg gtg aat ctg ggc gag ctg gtt gag tcc 107  
 Met Ala Lys Gly Leu Arg Val Asn Leu Gly Glu Leu Val Glu Ser  
 -30 -25 -20  
 atg cgt ttg tgc ttc ctc tca gtc cac ttt cgc tta cga tgg ggc gac 155  
 Met Arg Leu Cys Phe Leu Ser Val His Phe Arg Leu Arg Trp Gly Asp  
 -15 -10 -5 1  
 tct tgt cca tcg tca cct cac cgg gaa act ttt cct gcc ggg cca gtt 203  
 Ser Cys Pro Ser Ser Pro His Arg Glu Thr Phe Pro Ala Gly Pro Val  
 5 10 15  
 aat ggt ccc ctg tac cac ccc cgg g 228  
 Asn Gly Pro Leu Tyr His Pro Arg  
 20 25

<210> 381  
 <211> 300  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS

&lt;222&gt; 39..299

&lt;221&gt; sig\_peptide

&lt;222&gt; 39..89

&lt;223&gt; Von Heijne matrix

score 5.09999990463257

seq QLLVLFGSQTGTA/QD

&lt;400&gt; 381

agtttttagt ctcagaccag accaccgggc gcgccccg atg ccg agc ccg cag ctt 56  
 Met Pro Ser Pro Gln Leu

-15

ctg gtg ctc ttc ggc agc cag aca ggc acg gct cag gat gtg tcg gag 104  
 Leu Val Leu Phe Gly Ser Gln Thr Gly Thr Ala Gln Asp Val Ser Glu

-10

-5

1

5

aga ctg ggt cgc gag gcc cgg ggc cgg cgg ctt ggc tgc cgg gtg cag 152  
 Arg Leu Gly Arg Glu Ala Arg Gly Arg Arg Leu Gly Cys Arg Val Gln

10

15

20

gcc ctg gac tcc tac ccg gtg gtg aat ctg att aac gag ccc ctg gtg 200  
 Ala Leu Asp Ser Tyr Pro Val Val Asn Leu Ile Asn Glu Pro Leu Val

25

30

35

ata ttt gtt tgt gca act ayw ggc caa gga gac ccc cct gac aac atg 248  
 Ile Phe Val Cys Ala Thr Xaa Gly Gln Gly Asp Pro Pro Asp Asn Met

40

45

50

aag aac ttc tgg agg ttt ata ttc cgg aag aac ctg ccc tcc acc gcc 296  
 Lys Asn Phe Trp Arg Phe Ile Phe Arg Lys Asn Leu Pro Ser Thr Ala

55

60

65

cgg g 300

Arg

70

&lt;210&gt; 382

&lt;211&gt; 151

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 8..151

&lt;221&gt; sig\_peptide

&lt;222&gt; 8..130

&lt;223&gt; Von Heijne matrix

score 5.09999990463257

seq SFLFLACIFQGXS/XX

&lt;400&gt; 382

atacata atg tct tcc att ttg ggt gtc tca tcc tca tgg tgg tat tta 49  
 Met Ser Ser Ile Leu Gly Val Ser Ser Ser Trp Trp Tyr Leu

-40

-35

-30

tat tat ggc tat tgt ata ttt gtt aaa aag tgc tct ttt tgc agt ttc 97  
 Tyr Tyr Gly Tyr Cys Ile Phe Val Lys Lys Cys Ser Phe Cys Ser Phe

-25

-20

-15

ctg ttc ctt gcc tgt att ttt caa ggc tkt tck ckt kat wca aac aca 145  
 Leu Phe Leu Ala Cys Ile Phe Gln Gly Xaa Ser Xaa Xaa Xaa Asn Thr

-10

-5

1

5

caa agc 151

Gln Ser

&lt;210&gt; 383

&lt;211&gt; 255

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 101..253

&lt;221&gt; sig\_peptide

&lt;222&gt; 101..184

&lt;223&gt; Von Heijne matrix

score 5.09999990463257

seq CLCGSAPCLLCRC/CP

&lt;400&gt; 383

```

gcgtccggaa gtgtctcgca gatagtaa atctctcgaa aggcgagaaa gaagctgtct      60
ccatcttgct tgtatccgct gcwcttgtga cggtgtggag atg ggg agc gtc ctg      115
                                   Met Gly Ser Val Leu
                                   -25
ggg ctg tgc tcc atg gcg agc tgg ata cca tgt ttg tgt gga agt gcc      163
Gly Leu Cys Ser Met Ala Ser Trp Ile Pro Cys Leu Cys Gly Ser Ala
               -20               -15               -10
ccg tgt ttg cta tgc cga tgc tgt cct agt gga aac aac tcc act gta      211
Pro Cys Leu Leu Cys Arg Cys Cys Pro Ser Gly Asn Asn Ser Thr Val
               -5               1               5
act aga ttg atc tat gca ctt ttc ttg ctt gtt gga gta tgg gg          255
Thr Arg Leu Ile Tyr Ala Leu Phe Leu Leu Val Gly Val Trp
10               15               20

```

&lt;210&gt; 384

&lt;211&gt; 456

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 128..454

&lt;221&gt; sig\_peptide

&lt;222&gt; 128..265

&lt;223&gt; Von Heijne matrix

score 5.09999990463257

seq IGSCSVMSSGALC/VP

&lt;400&gt; 384

```

tacaactttg aaaagccctt cctctggctt gctagacagc tcattggaga ccgtaacttg      60
gaatttggtg ccattgcctgc tcttgccctca ccagagattg tcattggacc aaatttgga      120
gtgtagt atg agc wkr wtt agm agg ttg stt aga caa ctg ctc tcc cag      169
               Met Ser Xaa Xaa Xaa Arg Leu Xaa Arg Gln Leu Leu Ser Gln
               -45               -40               -35
rtg agg rwg atg acc tgt gag aat gaa gct gga gcc cag tgt car aag      217
Xaa Arg Xaa Met Thr Cys Glu Asn Glu Ala Gly Ala Gln Cys Gln Lys
               -30               -25               -20
tct agt ttt ata ggc agc tgt tct gtg atg tca agt ggt gca ctg tgt      265
Ser Ser Phe Ile Gly Ser Cys Ser Val Met Ser Ser Gly Ala Leu Cys
               -15               -10               -5
gtg cca ctt tat tat cta gct aag ggc aac atg tgc tcc atc tgt ggg      313
Val Pro Leu Tyr Tyr Leu Ala Lys Gly Asn Met Cys Ser Ile Cys Gly
1               5               10               15
atg ctg aag gag atg aat ggg ctt tgg agt gaa tgt gac agt tta aaa      361
Met Leu Lys Glu Met Asn Gly Leu Trp Ser Glu Cys Asp Ser Leu Lys
               20               25               30
aat acc ttc att gtt tgg rcc tgc ata ttt agc tgt ttg gga atg caa      409
Asn Thr Phe Ile Val Trp Xaa Cys Ile Phe Ser Cys Leu Gly Met Gln
               35               40               45
ttg awt tct tct kgr gtt tca aat gta aga ctg cta ctg tca cat ca      456

```

210

Leu Xaa Ser Ser Xaa Val Ser Asn Val Arg Leu Leu Leu Ser His  
 50 55 60

<210> 385  
 <211> 193  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 1..192

<221> sig\_peptide  
 <222> 1..78  
 <223> Von Heijne matrix  
 score 5.09999990463257  
 seq AFPFVCLTFCVGG/GP

<400> 385  
 atg cct cat cca ctg gct acc tct gcg ttt ctg cgt tcc gcc ttt cct 48  
 Met Pro His Pro Leu Ala Thr Ser Ala Phe Leu Arg Ser Ala Phe Pro  
 -25 -20 -15  
 ttt gtt tgt ctc acg ttt tgc gtg gga ggc ggt ccc ggg att tca ggg 96  
 Phe Val Cys Leu Thr Phe Cys Val Gly Gly Gly Pro Gly Ile Ser Gly  
 -10 -5 1 5  
 gtc tac cgg ctc ctt atg gcg aat gca acc cga aga gag agt gag gta 144  
 Val Tyr Arg Leu Leu Met Ala Asn Ala Thr Arg Arg Glu Ser Glu Val  
 10 15 20  
 agc ctc cgc ggg ttg ggc agg gac gga gag ggg gcc cgc gcg act cca g 193  
 Ser Leu Arg Gly Leu Gly Arg Asp Gly Glu Gly Ala Arg Ala Thr Pro  
 25 30 35

<210> 386  
 <211> 281  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 199..279

<221> sig\_peptide  
 <222> 199..267  
 <223> Von Heijne matrix  
 score 5.09999990463257  
 seq SLMVFLNLFFLNC/DP

<400> 386  
 tgtttatagg ttttaactct tatggttaga atggttgtga gtcatacgwg tgtcagacct 60  
 ctgctaattt cctcaggaca cattcccaga agtggaatta ccaagtcaaa gagcataaat 120  
 acttttagaga tacatgataa attgtgccag ctacctttcc aaaagagttg tactagttga 180  
 ggtttctgcc agcagtat atg aca gtt ggg ctc cat att tta aga gat tca 231  
 Met Thr Val Gly Leu His Ile Leu Arg Asp Ser  
 -20 -15  
 cta atg gtg ttt ctc aac ctt ttt ttt tta aac tgt gac cca cac agg 279  
 Leu Met Val Phe Leu Asn Leu Phe Phe Leu Asn Cys Asp Pro His Arg  
 -10 -5 1  
 gg 281

<210> 387  
 <211> 111  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 5..109

<221> sig\_peptide  
 <222> 5..67  
 <223> Von Heijne matrix  
 score 5.09999990463257  
 seq MFCVSLLLHHAYP/LP

<400> 387  
 cacc atg gta aga tgg gga cat ccc cct atg ttc tgt gtc tct ctc ctg 49  
 Met Val Arg Trp Gly His Pro Pro Met Phe Cys Val Ser Leu Leu  
 -20 -15 -10  
 ctc cac cat gct tat cct ttg cct tcc acc atg att gta agt ttc cca 97  
 Leu His His Ala Tyr Pro Leu Pro Ser Thr Met Ile Val Ser Phe Pro  
 -5 1 5 10  
 agg cct ccc ctg gg 111  
 Arg Pro Pro Leu

<210> 388  
 <211> 374  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 96..374

<221> sig\_peptide  
 <222> 96..173  
 <223> Von Heijne matrix  
 score 5.09999990463257  
 seq AMVCFGCPGGASS/RC

<221> misc\_feature  
 <222> 344  
 <223> n=a, g, c or t

<400> 388  
 ttttgccgc catgttttcg tcgcagtaac tgccttggtg tcagtagtca ttgccagttt 60  
 cgggcgttct ggacaattgg gatgctgcag agttc atg gct ggg gct gct cgt 113  
 Met Ala Gly Ala Ala Arg  
 -25  
 tgg gtg gga caa kaa tcc tct gca atg gtt tgt ttt ggc tgc cca gga 161  
 Trp Val Gly Gln Xaa Ser Ser Ala Met Val Cys Phe Gly Cys Pro Gly  
 -20 -15 -10 -5  
 ggt gcg tca agt cgc tgc cgc tcc cct cgt ggg cgt cag gcc tca aga 209  
 Gly Ala Ser Ser Arg Cys Arg Ser Pro Arg Gly Arg Gln Ala Ser Arg  
 1 5 10  
 gtt ccc cgc cta gaa aat gga gct cag cga gtc gtg cgt acc atg gtg 257  
 Val Pro Arg Leu Glu Asn Gly Ala Gln Arg Val Val Arg Thr Met Val  
 15 20 25  
 cac ctg gtt ttg cag cct aag cgg gtc act tta gtg cat cct cct cgc 305  
 His Leu Val Leu Gln Pro Lys Arg Val Thr Leu Val His Pro Pro Arg  
 30 35 40  
 gga ttg gag cct gtt tgc acc cct ata gcm vga atg arn ccc aag tca 353  
 Gly Leu Glu Pro Val Cys Thr Pro Ile Ala Xaa Met Xaa Pro Lys Ser  
 45 50 55 60  
 cac ggg ctc aga agt tct ttg 374  
 His Gly Leu Arg Ser Ser Leu

<210> 389  
 <211> 192  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 52..192

<221> sig\_peptide  
 <222> 52..153  
 <223> Von Heijne matrix  
 score 5.09999990463257  
 seq PXLXSLHGLLYG/SP

<400> 389  
 ggcagacttc aaccaggctg tgggaggaga gctcagtggg gcacagagaa g atg ggt 57  
 Met Gly  
 gtt gtc agt ggg ggt gtt ggt gac ttg acc aca aaa acc caa gag aat 105  
 Val Val Ser Gly Gly Val Gly Asp Leu Thr Thr Lys Thr Gln Glu Asn  
 -30 -25 -20  
 ggg ctc tta cca gvc cty ctc tcc wkc ctk cac gga ctg ctc tat ggc 153  
 Gly Leu Leu Pro Xaa Leu Leu Ser Xaa Leu His Gly Leu Leu Tyr Gly  
 -15 -10 -5  
 agc cct gat gca gar ctc acg ggc ccg gat ccc tgg gat 192  
 Ser Pro Asp Ala Glu Leu Thr Gly Pro Asp Pro Trp Asp  
 1 5 10

<210> 390  
 <211> 371  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 321..371

<221> sig\_peptide  
 <222> 321..365  
 <223> Von Heijne matrix  
 score 5.09999990463257  
 seq FLSXTCVLSCXRS/LS

<400> 390  
 tctgttcagg ttttgtatgt gttcatagta taatcttggg ttgtagggtg tgtgtatctg 60  
 ggaagaaact ttacaatctc taacaggcct ggaagggtcta atctataaaa gtatttcatt 120  
 gaccttgaag aagggtcaatt atttatataa gaaaataaac tcaacatttt atccataaaa 180  
 aatgtaattc cggaatttat gttagtaaaa ttataacact gataacataa aaagtgtat 240  
 taatccttaa gaaagagtta ccttttcttt tctatcttca tcacagctag cccagcttta 300  
 gtctatttca ttagcttctt atg ggc ttc ctc tca ckt aca tgc gtg ctc tct 353  
 Met Gly Phe Leu Ser Xaa Thr Cys Val Leu Ser  
 -15 -10 -5  
 tgc dtg cgc tcg ctc tct 371  
 Cys Xaa Arg Ser Leu Ser  
 1

<210> 391  
 <211> 328  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 184..327

<221> sig\_peptide  
 <222> 184..300  
 <223> Von Heijne matrix  
       score 5  
       seq LVCFFNSVSFLFG/VS

<400> 391  
 ccgttatgtg ttcagctcaa ttagattaat taccttcctc accaggagtc acaatgcttt       60  
 gcagttttatc tgcggtaact aaatgttagt tttgtaagta aaaggtagtg ttattgacct       120  
 cgaaagggct atagttcctt tgaacttaca gagaagagtt ccaaacaact atttctaacc       180  
 aag atg gaa tat ggg tca gca aaa ttg tct tca ggt aga gtt ttc tac       228  
       Met Glu Tyr Gly Ser Ala Lys Leu Ser Ser Gly Arg Val Phe Tyr  
                               -35                       -30                       -25  
 ttg cca aga gac ttt ggc att gag agg aga gtt ctt gtt tgt ttt ttt       276  
 Leu Pro Arg Asp Phe Gly Ile Glu Arg Arg Val Leu Val Cys Phe Phe  
                               -20                       -15                       -10  
 aac tct gta tca ttt ctg ttt ggt gtc tct ara aaa aaa tcc gra caa       324  
 Asn Ser Val Ser Phe Leu Phe Gly Val Ser Xaa Lys Lys Ser Xaa Gln  
                               -5                       1                       5  
 tgg g   328  
 Trp

<210> 392  
 <211> 303  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 252..302  
 <221> sig\_peptide  
 <222> 252..290  
 <223> Von Heijne matrix  
       score 5  
       seq MLSGLVLSWALA/YQ

<400> 392  
 tgaccttgta gcagttatct ttgttaaact ccttcatttc ttattttaaa taattaatta       60  
 attaatttag agacagggtc tcaactatgtc acccaggctg tagtgcagtg gtgcaatcat       120  
 ggctcactgt agccttgacc tcccaggctc aagcaatctt cctacctcag cctctcaggc       180  
 agctgggact acagaccac agcactacgc ctgacttatg attttatttt ttgtggagac       240  
 agggctcttac t atg ttg tct ggg ctt gtc tta aac tct tgg gcc tta gcc       290  
       Met Leu Ser Gly Leu Val Leu Asn Ser Trp Ala Leu Ala  
                               -10                       -5  
 tac caa cta gct g   303  
 Tyr Gln Leu Ala  
 1

<210> 393  
 <211> 366  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 298..366  
 <221> sig\_peptide

<222> 298..345  
 <223> Von Heijne matrix  
 score 5  
 seq VFFXGXSIILVLG/ST

<221> misc\_feature  
 <222> 265  
 <223> n=a, g, c or t

<400> 393  
 tttttcccccg cccctgagac cctgcagcac catctgtcat ggcggtctggg ctgttttggtt 60  
 tgagcgctcg ccgtcttttg gcggcagcgg cgacgcgagg gctcccggcc gcccgcgctcc 120  
 gctgggaatc tagctttctcc argamytgtg gtcgccccgt ccgctgtggc gggaaagcgg 180  
 tccccagaac cgaccacacc gtggcaagag gacccagaac ccgaggacga aaacttgtat 240  
 gagaagaasc cagactccca tggknatgac aaggaccccg ttttggacgt ctggaac 297  
 atg cga ctt gtc ttc ttc ktw ggc gks tcc atc atc ctg gtc ctt ggc 345  
 Met Arg Leu Val Phe Phe Xaa Gly Xaa Ser Ile Ile Leu Val Leu Gly  
 -15 -10 -5  
 agc acc ttt gkg gcc tat ctg 366  
 Ser Thr Phe Xaa Ala Tyr Leu  
 1 5

<210> 394  
 <211> 126  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 21..125

<221> sig\_peptide  
 <222> 21..68  
 <223> Von Heijne matrix  
 score 5  
 seq SDFLLLFVSLSL/PP

<400> 394  
 agcttggcat ataggctcaa atg tta tca tca gat ttt ttt ctc ctc ttt gtc 53  
 Met Leu Ser Ser Asp Phe Phe Leu Leu Phe Val  
 -15 -10  
 tct tta tct tta tct cca ttt cct ttt ttt ctt ttt cct ccc ctc ttt 101  
 Ser Leu Ser Leu Ser Pro Phe Pro Phe Phe Leu Phe Pro Pro Leu Phe  
 -5 1 5 10  
 tcc tgc ttt ctc tta ccc acc cgg g 126  
 Ser Cys Phe Leu Leu Pro Thr Arg  
 15

<210> 395  
 <211> 329  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 154..327

<221> sig\_peptide  
 <222> 154..195  
 <223> Von Heijne matrix  
 score 5  
 seq FIAALFTVAKIWN/QP



```

<400> 395
tgaaaatgta aattagtgc gttattatgg magtcagtat ggaacttcct caaaaaacta      60
acaataaaac tcccatatga tccagcaatc ctaccactgr atatttatcc aaaggaaagg      120
aagtcggtat atttaacagg catctgcacc ccc atg ttt att gca gca cta ttc      174
                               Met Phe Ile Ala Ala Leu Phe
                               -10

aca gta gcc aag ata tgg aat caa cct aaa tgt cca tca acg gat gaa      222
Thr Val Ala Lys Ile Trp Asn Gln Pro Lys Cys Pro Ser Thr Asp Glu
      -5              1              5
tgg ata aat aaa atg tgg tac ata tac aca atg gag tac tat cca gac      270
Trp Ile Asn Lys Met Trp Tyr Ile Tyr Thr Met Glu Tyr Tyr Pro Asp
10              15              20              25
ata aaa aag aat gga att ctg aca ttt aag gca aca agg atg aac cgg      318
Ile Lys Lys Asn Gly Ile Leu Thr Phe Lys Ala Thr Arg Met Asn Arg
      30              35              40

aag aca tta tg      329
Lys Thr Leu

<210> 396
<211> 99
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 5..97

<221> sig_peptide
<222> 5..49
<223> Von Heijne matrix
      score 5
      seq VCGCLCVWMCVCG/XV

<221> misc_feature
<222> 49
<223> n=a, g, c or t

<400> 396
gtat atg tgt gtg tgt ggg tgt tta tgt gtg tgg atg tgt gtg tgt ggn      49
      Met Cys Val Cys Gly Cys Leu Cys Val Trp Met Cys Val Cys Gly
      -15              -10              -5
wtt gtg tgt ata tac ata tgm gtg tat gtg tgt aca tgt gtg agg ggg      97
Xaa Val Cys Ile Tyr Ile Xaa Val Tyr Val Cys Thr Cys Val Arg Gly
1              5              10              15
ga      99

<210> 397
<211> 316
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 134..316

<221> sig_peptide
<222> 134..211
<223> Von Heijne matrix
      score 5
      seq LSLFVFFWLVGFS/FF

```

<221> misc\_feature  
 <222> 284..285  
 <223> n=a, g, c or t

<400> 397

```

ttgtgaactc ttctattatt attaagtgtt gtcaattgtc agcatccata ttctattccg      60
atgatgaata gaagcattat atttcagcat caaaatgcag ttggggtcgt aatgagcatc      120
attagggacc tta atg gga gtc aga act gta tgt cat ttt att cag gtt      169
          Met Gly Val Arg Thr Val Cys His Phe Ile Gln Val
          -25          -20          -15
ttt cta agt tta ttt gtg ttt ttt tgg tta gtt ggt ttt tct ttt ttc      217
Phe Leu Ser Leu Phe Val Phe Phe Trp Leu Val Gly Phe Ser Phe Phe
          -10          -5          1
ttt ttt tta cdb ttt tct acc aag cag gtg aga gtw gaa cag cat tgt      265
Phe Phe Leu Xaa Phe Ser Thr Lys Gln Val Arg Val Glu Gln His Cys
          5          10          15
gat ttt aaa agt aca cca nnd gta gag tct tcc agt acc gtt ggc cat      313
Asp Phe Lys Ser Thr Pro Xaa Val Glu Ser Ser Ser Thr Val Gly His
          20          25          30
gcc      316
Ala
35

```

<210> 398  
 <211> 251  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 63..251

<221> sig\_peptide  
 <222> 63..143  
 <223> Von Heijne matrix  
 score 5  
 seq LSCFYLLAIVSNA/VM

<400> 398

```

atgttgtagc ttctgtcata atttccttcc cttttaaggc tgaataattt tccattgtgt      60
at atg tac cat att ttg ttc atc cat tca ttc att gat aga tac ttg      107
  Met Tyr His Ile Leu Phe Ile His Ser Phe Ile Asp Arg Tyr Leu
          -25          -20          -15
agt tgc ttc tac ctt ttg gca att gtg agt aat gct gtt atg aac atg      155
Ser Cys Phe Tyr Leu Leu Ala Ile Val Ser Asn Ala Val Met Asn Met
          -10          -5          1
ggg gta caa atg tct gtt ttg agt cct tgt ttt gct ttc gtg cat tct      203
Gly Val Gln Met Ser Val Leu Ser Pro Cys Phe Ala Phe Val His Ser
          5          10          15          20
att aaa aat gtt aag gtt ctt tgc ttt tta ctt ttt ttt ctc ttt ggg      251
Ile Lys Asn Val Lys Val Leu Cys Phe Leu Leu Phe Phe Leu Phe Gly
          25          30          35

```

<210> 399  
 <211> 120  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 10..120

<221> sig\_peptide  
 <222> 10..75  
 <223> Von Heijne matrix  
       score 5  
       seq VQWLLVYSPSCAA/TI

<400> 399  
 tcatttacc atg cag ttc acc gtt tta atg tgt cca gtt cag tgg ttg tta     51  
           Met Gln Phe Thr Val Leu Met Cys Pro Val Gln Trp Leu Leu  
                   -20                   -15                   -10  
 gtg tat tca ccc agt tgt gca gcc acc atc aca gtc aat ttt aaa aca     99  
 Val Tyr Ser Pro Ser Cys Ala Ala Thr Ile Thr Val Asn Phe Lys Thr  
                   -5                   1                   5  
 ttt tca tca ccc caa acc ggg     120  
 Phe Ser Ser Pro Gln Thr Gly  
           10                   15

<210> 400  
 <211> 463  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 342..461

<221> sig\_peptide  
 <222> 342..452  
 <223> Von Heijne matrix  
       score 5  
       seq VSCLSAGLRVCCS/QR

<221> misc\_feature  
 <222> 246,260  
 <223> n=a, g, c or t

<400> 400  
 ctctgtcccc gcggctgggt ctctgtctgct ccggttcctg ggctcctaata tcttggtcca     60  
 gcttcttcca ggcacatcct cttctctgcc ctccgtccat tttggagccg gagatgggtg     120  
 gctkggggcc gcccagtag tgagacagtg gaagtaaacc ccatctgccg tccccgtgcg     180  
 tagagaaaaa cggtgaccgc gaggtctggg aggagagttg cctctgagga agaagggcac     240  
 agaganccaa aattagtttn gaaagcatcc tgatttggtg cccgaggcct ggaaagaaat     300  
 ggcggctggg gtgcggcgga ggtaggggag gaaaacgttg g atg aga agg gcc tgg     356  
   Met Arg Arg Ala Trp  
   -35  
 act cag gaa agg gaa ccg cgt ccg tgt gag ccc gct gag cgc gca gac     404  
 Thr Gln Glu Arg Glu Pro Arg Pro Cys Glu Pro Ala Glu Arg Ala Asp  
                   -30                   -25                   -20  
 cct gcc cct gtc tcc tgt ctg tct gca ggt ctg cgc gtc tgt tgt tcc     452  
 Pro Ala Pro Val Ser Cys Leu Ser Ala Gly Leu Arg Val Cys Cys Ser  
                   -15                   -10                   -5  
 cag cgc tct gc     463  
 Gln Arg Ser  
 1

<210> 401  
 <211> 206  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS

&lt;222&gt; 94..204

&lt;221&gt; sig\_peptide

&lt;222&gt; 94..168

&lt;223&gt; Von Heijne matrix

score 4.90000009536743

seq DFFICLLAICVSS/FE

&lt;400&gt; 401

tactgtttat tgattctttg attatggcca ttcttacagg agtaagggtgg tatcacactg 60

tggttttgat ttgcatttcc ctgatcatta gtg atg ttg cat ttg att tgc att 114

Met Leu His Leu Ile Cys Ile

-25 -20

tcc ctg atc gtt aat gat ttt ttc ata tgt ttg ttg gcc att tgc gta 162

Ser Leu Ile Val Asn Asp Phe Phe Ile Cys Leu Leu Ala Ile Cys Val

-15 -10 -5

tct tct ttt gag aat tgt cta ttt atg tcc tta gcc cac agt gg 206

Ser Ser Phe Glu Asn Cys Leu Phe Met Ser Leu Ala His Ser

1 5 10

&lt;210&gt; 402

&lt;211&gt; 330

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 42..329

&lt;221&gt; sig\_peptide

&lt;222&gt; 42..230

&lt;223&gt; Von Heijne matrix

score 4.90000009536743

seq VTSLANLIPPVKA/XP

&lt;400&gt; 402

acagggtccc actgcagtta ggagccggtg agtccgggtg g atg agg tca gag cgc 56

Met Arg Ser Glu Arg

-60

ccc atg gtg tgg tgc tgc ctc ttt gtc cgt tgc cag cga aaa cgg aaa 104

Pro Met Val Trp Cys Cys Leu Phe Val Arg Ser Gln Arg Lys Arg Lys

-55 -50 -45

cag agc acc caa gat gaa gat gct gtt agc ctt tgc agt ctc gac ata 152

Gln Ser Thr Gln Asp Glu Asp Ala Val Ser Leu Cys Ser Leu Asp Ile

-40 -35 -30

agt gag cct agt aat aaa cgg gtc aaa ccc ctt tcc cga gtc acg tgc 200

Ser Glu Pro Ser Asn Lys Arg Val Lys Pro Leu Ser Arg Val Thr Ser

-25 -20 -15

cta gca aac ctc atc ccg ccc gtg aag gcc ayg cca tta aag cgc ttc 248

Leu Ala Asn Leu Ile Pro Pro Val Lys Ala Xaa Pro Leu Lys Arg Phe

-10 -5 1 5

agt caa acc ctg cag cgc tcc att agc ttc cgc agt gag agt cgc cct 296

Ser Gln Thr Leu Gln Arg Ser Ile Ser Phe Arg Ser Glu Ser Arg Pro

10 15 20

gac atc ctc gcc ccc cga ccc tgg tcc aga aat g 330

Asp Ile Leu Ala Pro Arg Pro Trp Ser Arg Asn

25 30

&lt;210&gt; 403

&lt;211&gt; 311

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

<220>  
 <221> CDS  
 <222> 168..311

<221> sig\_peptide  
 <222> 168..227  
 <223> Von Heijne matrix  
 score 4.90000009536743  
 seq CILISTAFPSLLT/QI

<400> 403  
 tgagcagatg gtgccaggat ttaaacctat gtttatcaga tgcagatgac ccaaacagtg 60  
 gcttatctgt tggtaat tttt tagatc aagttaaaca taaatgactt tgcattactc 120  
 tttggtcact ttttcctagt catttcaaat agtctgtcctt atttctc atg gtt ttt 176  
 Met Val Phe  
 -20  
 tgg aca aaa ttt tgt att tta att agt aca gca ttt cct tct tta ttg 224  
 Trp Thr Lys Phe Cys Ile Leu Ile Ser Thr Ala Phe Pro Ser Leu Leu  
 -15 -10 -5  
 aca cag att att ttc cct aaa tct att aca ttt gct ttc cag ttt ttc 272  
 Thr Gln Ile Ile Phe Pro Lys Ser Ile Thr Phe Ala Phe Gln Phe Phe  
 1 5 10 15  
 tgg aac agg gaa aaa caa aaa aca aaa aca cca act ggg 311  
 Trp Asn Arg Glu Lys Gln Lys Thr Lys Thr Pro Thr Gly  
 20 25

<210> 404  
 <211> 274  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 80..274

<221> sig\_peptide  
 <222> 80..190  
 <223> Von Heijne matrix  
 score 4.90000009536743  
 seq MLIMLGIFNVHS/AV

<400> 404  
 ccctgcgagg gcatectggg ctttctccca ccgctttccg agcccgcttg cacctcggcg 60  
 atccccgact cccttcttt atg gcg tgc ctc ctg tgc tgt ggg ccg aag ctg 112  
 Met Ala Ser Leu Leu Cys Cys Gly Pro Lys Leu  
 -35 -30  
 gcc gcc tgc ggc atc gtc ctc agc gcc tgg gga gtg atc atg ttg ata 160  
 Ala Ala Cys Gly Ile Val Leu Ser Ala Trp Gly Val Ile Met Leu Ile  
 -25 -20 -15  
 atg ctc gga ata ttt ttc aat gtc cat tcc gct gtg ttg att gag gac 208  
 Met Leu Gly Ile Phe Phe Asn Val His Ser Ala Val Leu Ile Glu Asp  
 -10 -5 1 5  
 gtt ccc ttc acg gag aaa gat ttt gag aat ggc ccc cag aac ata tac 256  
 Val Pro Phe Thr Glu Lys Asp Phe Glu Asn Gly Pro Gln Asn Ile Tyr  
 10 15 20  
 aac ctt tac gag cat ggg 274  
 Asn Leu Tyr Glu His Gly  
 25

<210> 405  
 <211> 153  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 69..152

<221> sig\_peptide  
 <222> 69..116  
 <223> Von Heijne matrix  
 score 4.90000009536743  
 seq SALLLEXLQXAIP/RX

<400> 405  
 tttccctgc cctgtctct cattccctt cttctggagc atttcatcca cagaccctt 60  
 gcccaaga atg tct gtc tca gct ctg ctt cta gag mtc ctc caa gmt gcc 110  
 Met Ser Val Ser Ala Leu Leu Leu Glu Xaa Leu Gln Xaa Ala  
 -15 -10 -5  
 atc cct cgy mam acc tca ggc ttm caa gac ctg ccc aac tgg g 153  
 Ile Pro Arg Xaa Thr Ser Gly Xaa Gln Asp Leu Pro Asn Trp  
 1 5 10

<210> 406  
 <211> 206  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 57..206

<221> sig\_peptide  
 <222> 57..173  
 <223> Von Heijne matrix  
 score 4.90000009536743  
 seq VIAIVSFTTLCSS/LY

<400> 406  
 aaataaaaaa tattaaaaaa taatctcatc tttgatttta gatttagggg gtgtgc atg 59  
 Met  
 cag gct tgt tat atg ggt atg tgg tat act gcc gag gct tgg ggt acg 107  
 Gln Ala Cys Tyr Met Gly Met Trp Tyr Thr Ala Glu Ala Trp Gly Thr  
 -35 -30 -25  
 att gag tcc ctc acc cag gta gtg agc gta atc gca ata gtt agt ttt 155  
 Ile Glu Ser Leu Thr Gln Val Val Ser Val Ile Ala Ile Val Ser Phe  
 -20 -15 -10  
 aca acc ctg tgc tcc tct ctg tat tcc ccc caa gta gtc ccc agt gtt 203  
 Thr Thr Leu Cys Ser Ser Leu Tyr Ser Pro Gln Val Val Pro Ser Val  
 -5 1 5 10  
 ggg 206  
 Gly

<210> 407  
 <211> 479  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> 277..477

<221> sig\_peptide  
 <222> 277..462  
 <223> Von Heijne matrix  
 score 4.90000009536743

seq PLAACPLLLPIFS/HA

&lt;221&gt; misc\_feature

&lt;222&gt; 22

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 407

```

aatggattga gatggggaag anaaaaagcc ccaaattcat gaaatgtagc tgckacagtc      60
cccacctcct tagctgtccc caaaacctaa gcaggtaatc ataacttcca ttctgtgctc      120
accttacctc tgctggcacc tttttggaca gggttctcta cttggcgagg tgacccaaat      180
cttcattcct gcagggtctg agtcctmrgc cgctgcgatc gtttgaacat tgtttgctcc      240
cacmraaact catcttgagg cttgggtcccc actgta atg atg ttg aga ggt ggc      294
                               Met Met Leu Arg Gly Gly
                               -60

ggg aca ttt aag grg tgt ttg agt cat gag gga tcc agc ttc acg aag      342
Gly Thr Phe Lys Xaa Cys Leu Ser His Glu Gly Ser Ser Phe Thr Lys
-55                               -50                               -45

gga tta gcg cag gag tgc gtg agt rct tct tgt ggg act cga ttg att      390
Gly Leu Ala Gln Glu Cys Val Ser Xaa Ser Cys Gly Thr Arg Leu Ile
-40                               -35                               -30                               -25

act gca gtw gcc agt kgt tac aaa gca agg ctg cct ctg gcc gcg tgc      438
Thr Ala Val Ala Ser Xaa Tyr Lys Ala Arg Leu Pro Leu Ala Ala Cys
-20                               -15                               -10

ccd ctt ctg ctt cct att ttc tcc cat gct aga agc agc ac      479
Pro Leu Leu Leu Pro Ile Phe Ser His Ala Arg Ser Ser
-5                               1                               5

```

&lt;210&gt; 408

&lt;211&gt; 289

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 84..287

&lt;221&gt; sig\_peptide

&lt;222&gt; 84..203

&lt;223&gt; Von Heijne matrix

score 4.90000009536743

seq SLKICGLVFGILA/LT

&lt;400&gt; 408

```

agccgactca cttgcaactc cacctcagca gtggctcttc agtcctctca aagcaaggaa      60
agagtactgt gtgctgagag acc atg gca aag aat cct cca gag aat tgt gaa      113
                               Met Ala Lys Asn Pro Pro Glu Asn Cys Glu
                               -40                               -35

gac tgt cac att cta aat gca gaa gct ttt aaa tcc aag aaa ata tgt      161
Asp Cys His Ile Leu Asn Ala Glu Ala Phe Lys Ser Lys Lys Ile Cys
-30                               -25                               -20                               -15

aaa tca ctt aag att tgt gga ctg gtg ttt ggt atc ctg gcc cta act      209
Lys Ser Leu Lys Ile Cys Gly Leu Val Phe Gly Ile Leu Ala Leu Thr
-10                               -5                               1

cta att gtc ctg ttt tgg ggg agc aag cac ttc tgg ccg gag gta ccc      257
Leu Ile Val Leu Phe Trp Gly Ser Lys His Phe Trp Pro Glu Val Pro
5                               10                               15

aaa aaa gcc tat gac atg gag cac act acg gg      289
Lys Lys Ala Tyr Asp Met Glu His Thr Thr
20                               25

```

&lt;210&gt; 409

&lt;211&gt; 341